GENOME-WIDE STUDY OF EMPHYSEMA PROGRESSION AND GENE-BY-SMOKING INTERACTION OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE

by
Woori Kim, MPH

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
January 2020

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is a progressive respiratory disease characterized by airflow limitation, an impairment of lung function. Emphysema is a major pathological change of COPD and characterized by the destruction of lung parenchyma. COPD is the third leading cause of death worldwide. The primary environmental risk factor for COPD is cigarette smoking, but individuals vary in their susceptibility to the effects of cigarette smoke and only a minority of smokers will ever develop COPD. This indicates other factors besides smoking are involved in the development and progression of COPD, and these may interact with smoking. This dissertation investigates the genetic determinants of emphysema progression quantified by computed tomography (CT) imaging and studies gene-by-smoking interaction on COPD risk 1) at the genome-wide scale and 2) using a polygenic risk score approach. This dissertation uses data from the COPDGene and ECLIPSE study, a longitudinal cohort of current and former smokers measuring chest CT scan, and the UK Biobank, a large-scale cohort recruiting over 500,000 volunteers throughout the United Kingdom.

Thesis committee

Terri H. Beaty, PhD (advisor)
Professor of Epidemiology

Christine Ladd-Acosta, PhD
Assistant Professor of Epidemiology

Nilanjan Chatterjee, PhD
Professor of Biostatistics

Thesis readers

Terri H. Beaty, PhD (advisor)
Professor of Epidemiology

Nilanjan Chatterjee, PhD
Professor of Biostatistics

Nadia N. Hansel, MD
Professor of Medicine

Miranda R. Jones, PhD
Assistant Professor of Epidemiology

Acknowledgements

My dissertation would not be possible without the help of many people. First and foremost, I would like to thank my advisor Dr. Terri Beaty for all her guidance and support. Thank you for encouraging me to get involved in many research experiences, and thanks for all the opportunities you have given me. Thank you for providing me with the opportunity to have a research experience in Channing division at Harvard during my PhD program. I would not have made it through the PhD program without you.

I would also like to thank my thesis committee, Drs. Christine Ladd-Acosta and Nilanjan Chatterjee. Thank you for all the time you spent in my committee meetings, for sharing your expertise with me, and for the encouragement along the way. I truly appreciate it.

This dissertation would not be possible without the help of my COPDGene collaborators, especially Michael Cho, Ed Silverman, Phuwanat, Brian and Dmitry, and other Channing investigators. Thanks for the opportunity to work on this project, and thanks for all of your input along the way.

I sincerely thank all the faculty members and students in the genetic epidemiology department. I have felt supported by the whole department and grateful to have many teachers along the way. A special thanks to Drs. Priya Duggal and Debashree Ray for your encouragement, kindness and enthusiasm. Thanks for your teaching and mentoring along the way. I also would like to thank to the epidemiology department staffs, particularly Fran and Matt for all the help about PhD program and funding source.

Thanks to all those who studied together, worked in W6517 and shared a wonderful memory at Bar Harbor Jackson summer course including Cristian, Kruthika, Yanzi, Nora, Steven, Dylan

and doctoral student cohort. I would not have made it through this process without your support, advice, and friendship.

I would like to thank my wonderful friends from Hopkins, Boston and Korea. You made my PhD program fun and meaningful.

I would like to thank my family for their heartful and continuing support. To my parents, and to my sisters, Hannah and Cheonga, I could not have done it without you. Last but not least, thank you, God!

Dedication

This dissertation is dedicated to my parents, Jaegon Kim and Sangsook Kim, for their eternal love, trust and support.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction and literature review

Chapter 1. Introduction and literature review

Chronic Obstructive Pulmonary Disease (COPD) is a progressive respiratory disease characterized by airflow limitation, an impairment of lung function. COPD patients present persistent respiratory symptoms such as dyspnea, chronic cough, and chronic sputum production. COPD is caused by a mixture of two subtypes, small airways disease and parenchymal destruction (emphysema) [1]. Emphysema is a major pathological change of COPD. It is characterized by the abnormal and permanent enlargement of distal airspaces along with destruction of lung parenchyma. As COPD is a highly heterogeneous disease, COPD patients show marked variability in their lung abnormality and their respiratory symptoms.

COPD burden

The burden of COPD has increased in the last two decades. The Global Burden of Disease Study reported a prevalence of 251 million cases of COPD globally in 2016 [2]. It is estimated that 5% (estimated 3.17 million deaths) of all deaths was caused by COPD in 2015. In the United States (U.S.), COPD became the third leading cause of death in 2014 [3]. COPD affects 5-10% of the U.S. population with an estimated 24 million affected individuals [4].

COPD results in substantial economic and social burden. COPD is the second leading cause of reduced Disability-Adjusted Life Year (DALY) in U.S. The estimated direct and indirect costs of COPD are $32 billion and $20.4 billion, respectively, in U.S. [5]. With the aging population and cumulative exposures to COPD risk factors, the burden of COPD is projected to increase over the next 30 years globally. By 2030, there may be over 4.5 million deaths annually from COPD and its related conditions [2].

Measurement of COPD

Spirometry measure of lung function

Spirometry measures two primary lung functions to assess an airflow limitation; forced expiratory volume in the first second (FEV_1) is the volume of air forcibly exhaled in the first second and forced vital capacity (FVC) is the volume of air that can be maximally forcibly exhaled. The presence of a post-bronchodilator FEV_1/FVC ratio less than 0.70 confirms the presence of airflow obstruction, and thus defines clinical COPD. The predicted post-bronchodilator FEV_1 grades disease severity of COPD into mild, moderate, severe and very severe according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [1].

Spirometry is a noninvasive and readily available tool to diagnose and assess COPD in the clinical setting. However, spirometry alone cannot distinguish two key subtypes of COPD: emphysema and small airways disease. COPD patients diagnosed based on spirometry alone still have significant differences in their underlying lung abnormality.

CT measures

Computed Tomography (CT) imaging can help explain some heterogeneity in COPD. CT imaging analysis can detect emphysema visually and quantify the level of emphysema. Previous studies have shown quantitative emphysema measured by CT imaging are correlated with airflow limitation measured by FEV_1 and FVC [6–9]. Quantitative CT measures of emphysema provide a quantitative assessment of emphysema progression over time.

Inherently, CT measures have a large variability. In the research setting, two common CT measures have shown meaningful results: percent emphysema and adjusted lung density [10]. The percent emphysema is the percentage of low-attenuation areas (i.e. those less than -950

Hounsfield Units (HU) on inspiratory CT). The adjusted lung density is the 15th percentile of the HU distribution adjusted for predicted total lung volume (also measured by CT) on inspiratory CT. Higher percent emphysema and lower adjusted lung density indicate greater severity of emphysema.

Etiology of COPD

COPD is a complex disease influenced by multiple environmental and genetic factors. The primary environmental risk factor for COPD is cigarette smoking, but individuals vary in their susceptibility to the effects of cigarette smoke and only a minority of smokers will ever develop COPD. This indicates other factors besides smoking are involved in the development and progression of COPD, and these may interact with smoking.

Environmental factors

Cigarette smoking

Cigarette smoking is considered a primary environmental cause of COPD. In a 25-year prospective cohort study of 8,045 individuals, current smokers were 6.3 times more likely to develop COPD than never smokers [11]. Approximately 80% of COPD deaths are caused by smoking, and smokers are 12-13 times more likely to die from COPD than never smokers [12]. However, a substantial proportion of COPD cases cannot be attributed to smoking alone. The majority of population-attributable fraction (PAF) for smoking was less than 80% [13]. The prevalence of COPD among never smokers ranges from 3 to 15% [14,15]. These studies indicate other risk factors, besides cigarette smoking, are important in the development of COPD.

Sex

There are distinct sex differences in COPD prevalence. It has been known that COPD primarily affects men, because men have smoked more than women worldwide [14]. However, the prevalence of COPD has been increasing in women and the mortality due to COPD is now similar between men and women in U.S. [16,17]. There is evidence suggesting there may be enhanced susceptibility to cigarette smoking in women compared to men, resulting in sex differences in COPD. In the National Emphysema Treatment Trial (NETT) study, females had fewer pack-years of smoking history than men but had similar severity of COPD [18]. In a systematic meta-analysis by Gan et al., females had a faster annual rate of decline in lung function (measured as FEV_1) than males of similar smoking levels [19]. The effect of sex on COPD development and progression is likely due to a combination of environmental and biological factors, however the exact mechanism resulting in this disease disparity remains unresolved.

Race

There are also differences in lung function between European Americans (EAs) and African Americans (AAs). According to the Centers for Disease Control and Prevention (CDC), the age-adjusted mortality rate from COPD is 78.1 per 100,000 in EAs and 63.9 per 100,000 in AAs [20]. However, mortality from COPD is increasing more rapidly among AAs [21] and there is evidence that AAs may be more susceptible to the harmful effects of tobacco smoke [22,23]. The potential impact of these racial differences in COPD development needs further investigation.

Age

Aging is associated with a progressive decline in lung function even among non-smoking adults [24]. It remains unclear if this change results from a normal aging process or if age reflects the sum of cumulative exposures throughout life leading to lung function decline.

Other environmental factors

A number of additional environmental factors are associated with risk of COPD including chronic asthma [25], outdoor air pollution [26], secondhand smoke exposure [27], biomass smoke [28], occupational exposure [29,30], diet [31] and tuberculosis [32]. However, the specific causal relationships among these factors and their individual or combined effects on risk to COPD remain unresolved.

Genetic risk factors

There is evidence of genetic influences on COPD and its related traits. The estimated heritability of spirometric measures of lung function (FEV_1 and FVC which are used to define COPD) was approximately 40% [33]. The longitudinal decline in lung function is reported to be moderately heritable; the estimated heritability of FEV_1, FVC and FEV_1/FVC was 0.05, 0.18 and 0.13, respectively [34]. The heritability of COPD status and quantitative emphysema measured by CT scan was estimated as 37% and 25%, respectively [35].

Alpha-1 antitrypsin deficiency

An established gene influencing risk of COPD is *SERPINA1* which encodes serine protease alpha-1-antitrypsin (AAT) protein. Mutations in *SERPINA1* leads to the hereditary AAT deficiency. Homozygosity for the PI*Z allele is the most common cause of AAT deficiency. The AAT deficiency follows a Mendelian pattern of inheritance, but there is marked variability in the development and severity of COPD among PI*ZZ individuals. This suggests a potential *SERPINA1*-by-environment interaction. However, this gene accounts for only about 1-3% of all COPD cases due to the rarity in the PI*Z allele.

Genome-wide association studies

A number of studies have attempted to identify genetic determinants of COPD and its subtypes including emphysema quantitatively measured by CT scan using genome-wide association studies (GWAS).

Emphysema. Multiple GWASs have been identified for cross-sectional emphysema measures based on CT imaging [36–40]. Using two quantitative CT measures, percent emphysema and adjusted lung density, 7 genetic regions have been identified as associated with cross-sectional quantitative emphysema (*DLC1, SERPINA1, HHIP, CHRNA3, AGER, SNRPF, BICD1)* [36,37,41]. However, to date, little is known about genetic determinants of progression of emphysema. As emphysema is a progressive disease, genetic factors may well be involved in the progression of emphysema.

COPD. The most recent and largest GWAS in 35,735 cases and 222,076 controls from the UK Biobank and additional studies from the International COPD Genetics Consortium (ICGC) identified 82 distinct loci associated with risk of COPD [42]. Forty-seven loci had been previously identified in GWAS of COPD or population-based measures of lung function [43,44]. Thirty-five loci were newly identified in this GWAS study. However, the identified variants at these distinct genetic regions explained less than 10% of the phenotypic variability on the liability scale.

In an early GWAS of COPD, two genetic loci, containing *HHIP* on chromosome 4q31 and *CHRNA3/A5/B4* on 15q25, were identified [45]. *HHIP* (the hedgehog interacting protein) gene encodes a regulatory protein in the hedgehog signaling pathway and may influence fetal lung development. The 15q25 region has been intriguing but challenging to understand. This genomic region contains a gene cluster *CHRNA3/A5/B4* related to smoking behavior and

nicotine addiction, but it also appears to contain COPD genetic determinants unrelated to nicotine addiction – for example, *IREB2*. Subsequent work also identified *FAM13A* on chromosome 4q22 as a genetic risk factor for COPD [46]. Of the genome-wide significant association regions for COPD, 7 genes (*HHIP, FAM13A, IREB2, AGER, MMP1, MMP12,* and *SFTPD*) point to potentially relevant biological processes involved in COPD pathogenesis based on a knock-out or transgenic murine model [47].

Lung function. Concurrently, the most recent and largest GWAS identified 279 distinct signals of lung function (used to define clinical COPD) in 400,102 individuals of European ancestry from the UK Biobank and SpiroMeta Consortium [48]. Of these 279 lung function signals, 139 signals were newly identified. The study reported an enrichment for genes involved in ciliogenesis (including *KIAA0753*, *CDK2* and *CEP72*). New signals implicating *ITGAV* and *GDF5*, as well as stronger evidence for *TGFB2* and *MFAP2* provided new genetic support for the importance of elastic fiber pathways in controlling quantitative measures of lung function and risk to COPD.

Gene-by-Environment Interaction

Broadly, gene-by-environment interaction is defined as that the relationship between phenotype and genotype is modified by some environmental risk factors. Gene-by-environment interaction is an important component to understand the disease mechanism. There are several important applications for evidence of gene-by-environment interactions: 1) it helps to improve the understanding of the biology of the disease, 2) it can improve predictions of risk to disease and 3) it can guide development of targeted interventions (a clear public health application).

The adverse effects of smoking on risk of COPD may differ by an individual's genetic susceptibility, which raises the possibility of detectable gene-by-smoking interactions using genetic data on large samples. However, little is known about gene-by-smoking interactions on COPD risk. One robustly replicated interaction is the PI*Z allele (*SERPINA1*)-by-smoking interaction on spirometric measures of lung function [49,50].

Genome-wide approach

Gene-by-environment interaction studies have been commonly investigated using a single marker in candidate genes or at the genome-wide scale. A major challenge to detect and replicate gene-by-environment interactions is the much larger sample size required compared to conventional GWAS to detect marginal effects of genes [51]. To date, genome-wide gene-by-smoking interaction studies have focused on quantitative measures of lung function [52,53]. While spirometric measures of lung function are used to diagnose COPD, no genome-wide studies have investigated gene-by-smoking interaction on risk of COPD itself.

Polygenic risk score approach

While the genome-wide approach for identifying potential gene-by-environment interactions can help to discover novel gene-by-environment interactions, it is quite challenging to detect and replicate such interactions due to a very small effect of each single marker on disease risk in general. Several studies have attempted to utilize polygenic risk scores (PRS) to explore gene-by-environment interactions and found some intriguing findings, such as PRS-by-physical activity or PRS-by-sugar consumption interaction on BMI [54–57]. PRS is a sum of the number of risk alleles for each SNP weighted by the estimated effect size reported from GWASs. An empirical study suggested PRS-based tests for interaction can outperform other genome-wide

approaches to detect gene-by-environment interactions, if the interaction effects of each single SNP tend to go in the same direction [58].

Recent large-scale GWASs of lung function and COPD have identified multiple genetic loci (279 signals for lung function and 82 signals for risk to COPD). With these well-established SNPs, one can explore the aggregated effect on risk to COPD by constructing PRS based on these recognized genetic risk factors and then testing for interaction between PRS and smoking on risk of COPD. Previous study investigated the potential interaction between PRS of 26 lung function SNPs and smoking on spirometric measures of lung function [58], highlighting the benefit of using PRS for identifying interactions missed when studying individual SNPs only.

Motivation and specific aims for this study

Little is known about genetic determinants of progression of emphysema. Identification of genetic factors for emphysema progression may identify different or more specific pathways than an analysis of overall lung function decline, expand our understanding of the genetics of COPD, and perhaps contribute to development of new drug therapies to slow the loss of lung density. Despite a well-recognized relationship between cigarette smoking and risk to COPD, smoking effects likely vary among individuals due to their genetic makeup, suggesting a potential gene-by-smoking interaction. However, to date, gene-by-smoking interactions on risk to COPD remain largely unknown. We seek to address these issues through the following specific aims:

Aim 1. To examine genetic variants for association with change in quantitative emphysema measured by CT imaging from two longitudinal cohort studies:

COPDGene and Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE)

Aim 2. To identify novel genetic variants for risk of COPD while accounting for potential smoking interactions and assess the impact of gene-by-smoking interactions on risk of COPD at known COPD and lung function GWAS loci

Aim 3. To examine whether the effect of smoking on COPD risk is modified by a polygenic risk score of lung function

Study description

COPDGene

The Genetic Epidemiology of COPD (COPDGene) study is a multicenter longitudinal cohort enrolling Non-Hispanic Whites (NHW) and African Americans (AA) from multiple sites in the US [59]. The COPDGene recruited subjects between ages of 45 and 80 years with a minimum of 10 pack-years smoking history at baseline and conducted volumetric inspiratory CT scans of the chest. Approximately 5 years later, subjects were asked to return to repeat the chest CT scan and detailed questionnaires. CT measures were available on 5,093 subjects COPDGene subjects at both visit 1 and visit 2. Subjects were genotyped on the Illumina Human Omni Express array (San Diego, CA). Genotyping quality control (QC) was performed following previously described guidelines to remove low quality subjects and markers [46,60]. Unobserved genotypes were imputed using Michigan Imputation Server with the Haplotype Reference Consortium (HRC) panel [61].

ECLIPSE

The Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE) is a 3-year longitudinal study. The ECLIPSE recruited COPD cases and controls aged 40 to 75 years with a minimum of 10 pack-years smoking history at baseline. Subjects included in this dissertation are self-reported Whites from both American and European sites. Chest CT scan was conducted at baseline, after 1 year and after 3 years in the ECLIPSE study. CT measures were available on 1,871 subjects who had at least two visits including the baseline. subjects were genotyped on the Illumina HumanHap 550 V3 array (San Diego, CA). Genotyping QC was performed following previously described guidelines to remove low quality subjects and markers [46,60]. Unobserved genotypes were imputed using Michigan Imputation Server with the HRC panel [61].

UK Biobank

The UK Biobank (UKB) is a population-based large cohort of volunteers where over 500,000 individuals aged 40-69 were originally recruited [62]. Extensive baseline questionnaire data, physiologic measures including spirometric measures of lung function, and biologic specimens (urine and blood samples) have been obtained. To determine lung function, measures of FEV_1 and FVC were derived from the spirometry volume-time series data, subjected to additional quality control based on American Thoracic Society (ATS) / European Respiratory Society (ERS) criteria [48,63]. Genotyping was performed by using the Axiom UK BiLEVE array and the Axiom Biobank array (Affymetrix) and imputed to the HRC version 1.1 panel [61].

Summary

Many recent advances have been made in our understanding of the genetics of COPD. However, specific genetic determinants and potential gene-by-environment interactions

influencing risk to COPD remain still largely unknown. The overall goal of this dissertation is to understand the genetic control of the progression of emphysema over time and to test for gene-by-smoking interactions on risk to COPD. Improved understanding of the genetics and gene-by-environment interactions influencing risk to COPD can aid in the discovery of novel therapeutics and a more tailored treatment, and ultimately reduce the burden of COPD.

References

1. GOLD. *Global Initiative for Chronic Obstructive Lung Disease (GOLD)*.; 2017. http://goldcopd.org.
2. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet (London, England)*. 2017;390(10100):1211-1259. doi:10.1016/S0140-6736(17)32154-2
3. Statistics NC for H. National Vital Statisitcs Reports, Volume 65, Number 4, (06/30/2016). 2016:1-122. https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_04.pdf. Accessed November 14, 2017.
4. Mannino DM, Homa DM, Akinbami LJ, Ford ES, Redd SC. Chronic obstructive pulmonary disease surveillance--United States, 1971-2000. *MMWR Surveill Summ*. 2002;51(6):1-16. http://www.ncbi.nlm.nih.gov/pubmed/12198919. Accessed December 2, 2017.
5. National Heart, Lung and BI. *2012 NHLBI Morbidity and Mortality Chart Book*.; 2012. http://www.nhlbi.nih.gov/research/reports/2012-mortality-chart-book.
6. Gould GA, Redpath AT, Ryan M, et al. Lung CT density correlates with measurements of airflow limitation and the diffusing capacity. *Eur Respir J*. 1991;4(2):141-146. http://www.ncbi.nlm.nih.gov/pubmed/2044729. Accessed November 14, 2017.
7. Nakano Y, Muro S, Sakai H, et al. Computed tomographic measurements of airway dimensions and emphysema in smokers. Correlation with lung function. *Am J Respir Crit Care Med*. 2000;162(3 Pt 1):1102-1108. doi:10.1164/ajrccm.162.3.9907120
8. Washko GR, Criner GJ, Mohsenifar Z, et al. Computed tomographic-based quantification of emphysema and correlation to pulmonary function and mechanics. *COPD*.

2008;5(3):177-186. doi:10.1080/15412550802093025

9. Schroeder JD, McKenzie AS, Zach JA, et al. Relationships between airflow obstruction and quantitative CT measurements of emphysema, air trapping, and airways in subjects with and without chronic obstructive pulmonary disease. *AJR Am J Roentgenol*. 2013;201(3):W460-70. doi:10.2214/AJR.12.10102
10. Lynch DA, Al-Qaisi MA. Quantitative Computed Tomography in Chronic Obstructive Pulmonary Disease. *J Thorac Imaging*. 2013;28(5):284-290. doi:10.1097/RTI.0b013e318298733c
11. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2095-2128. doi:10.1016/S0140-6736(12)61728-0
12. U.S. Department of Health and Human Services. *The Health Consequences of Smoking: A Report of the Surgeon General.* Vol 2013.; 2004. doi:10.1002/yd.20075
13. Eisner MD, Anthonisen N, Coultas D, et al. An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2010;182(5):693-718. doi:10.1164/rccm.200811-1757ST
14. Buist AS, McBurnie MA, Vollmer WM, et al. International variation in the prevalence of COPD (The BOLD Study): a population-based prevalence study. *Lancet*. 2007. doi:10.1016/S0140-6736(07)61377-4
15. Menezes AMB, Perez-Padilla R, Jardim JRB, et al. Chronic obstructive pulmonary disease in five Latin American cities (the PLATINO study): A prevalence study. *Lancet*. 2005. doi:10.1016/S0140-6736(05)67632-5
16. S. A, E. D-G, D.M. M. COPD and gender differences: An update. *Transl Res*. 2013.
17. Cohen SBZ, Paré PD, Man SFP, Sin DD. The growing burden of chronic obstructive pulmonary disease and lung cancer in women: Examining sex differences in cigarette smoke metabolism. *Am J Respir Crit Care Med*. 2007. doi:10.1164/rccm.200611-1655PP
18. Martinez FJ, Curtis JL, Sciurba F, et al. Sex differences in severe pulmonary emphysema. *Am J Respir Crit Care Med*. 2007. doi:10.1164/rccm.200606-828OC
19. Gan WQ, Man SFP, Postma DS, Camp P, Sin DD. Female smokers beyond the perimenopausal period are at increased risk of chronic pulmonary disease: A systematic review and meta-analysis. *Respir Res*. 2006. doi:10.1186/1465-9921-7-52

20. Centers for Disease Control and Prevention. *CDC Wonder*.; 2016.
21. Dransfield MT, Bailey WC. COPD: Racial Disparities in Susceptibility, Treatment, and Outcomes. *Clin Chest Med*. 2006. doi:10.1016/j.ccm.2006.04.005
22. Hansel NN, Washko GR, Foreman MG, et al. Racial Differences in CT Phenotypes in COPD. *COPD J Chronic Obstr Pulm Dis*. 2013;10(1):20-27. doi:10.3109/15412555.2012.727921
23. Wise RA. Changing smoking patterns and mortality from chronic obstructive pulmonary disease. *Prev Med (Baltim)*. 1997. doi:10.1006/pmed.1997.0181
24. Ware JH, Dockery DW, Louis TA, Xu X, Ferris BG, Speizer FE. Longitudinal and cross-sectional estimates of pulmonary function decline in never-smoking adults. *Am J Epidemiol*. 1990. doi:10.1093/oxfordjournals.aje.a115710
25. Braman SS, Kaemmerlen JT, Davis SM. Asthma in the elderly: A comparison between patients with recently acquired and long-standing disease. *Am Rev Respir Dis*. 1991. doi:10.1164/ajrccm/143.2.336
26. Sunyer J. Urban air pollution and chronic obstructive pulmonary disease: A review. *Eur Respir J*. 2001. doi:10.1183/09031936.01.17510240
27. Eisner MD, Balmes J, Katz PP, Trupin L, Yelin EH, Blanc PD. Lifetime environmental tobacco smoke exposure and the risk of chronic obstructive pulmonary disease. *Environ Heal*. 2005;4(1):7. doi:10.1186/1476-069X-4-7
28. Smith KR. *Biofuels, Air Pollution, and Health*.; 1988. doi:10.1007/978-1-4613-0891-1
29. Love RG, Miller BG. Longitudinal study of lung function in coal-miners. *Thorax*. 1982. doi:10.1136/thx.37.3.193
30. Seixas NS, Robins TG, Attfield MD, Moulton LH. Longitudinal and cross sectional analyses of exposure to coal mine dust and pulmonary function in new miners. *Br J Ind Med*. 1993. doi:10.1136/oem.50.10.929
31. Romieu I, Trenga C. Diet and Obstructive Lung Diseases. *Epidemiol Rev*. 2001;23(2):268-287. doi:10.1093/oxfordjournals.epirev.a000806
32. Hnizdo E, Singh T, Churchyard G. Chronic pulmonary function impairment caused by initial and recurrent pulmonary tuberculosis following treatment. *Thorax*. 2000. doi:10.1136/thorax.55.1.32
33. Palmer LJ, Knuiman MW, Divitini ML, et al. Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. *Eur Respir J*. 2001;17(4):696-702. http://www.ncbi.nlm.nih.gov/pubmed/11401066. Accessed November 19, 2017.

34. Gottlieb DJ, Wilk JB, Harmon M, et al. Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med*. 2001;164(9):1655-1659. doi:10.1164/ajrccm.164.9.2010122
35. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med*. 2013;188(8):941-947. doi:10.1164/rccm.201302-0263OC
36. Kong X, Cho MH, Anderson W, et al. Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. *Am J Respir Crit Care Med*. 2011;183(1):43-49. doi:10.1164/rccm.201004-0541OC
37. Manichaikul A, Hoffman EA, Smolonska J, et al. Genome-wide study of percent emphysema on computed tomography in the general population. The Multi-Ethnic Study of Atherosclerosis Lung/SNP Health Association Resource Study. *Am J Respir Crit Care Med*. 2014;189(4):408-418. doi:10.1164/rccm.201306-1061OC
38. Castaldi PJ, Cho MH, San José Estépar R, et al. Genome-wide association identifies regulatory Loci associated with distinct local histogram emphysema patterns. *Am J Respir Crit Care Med*. 2014;190(4):399-409. doi:10.1164/rccm.201403-0569OC
39. Cho MH, Castaldi PJ, Hersh CP, et al. A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. *Am J Respir Crit Care Med*. 2015;192(5):559-569. doi:10.1164/rccm.201501-0148OC
40. Boueiz A, Lutz SM, Cho MH, et al. Genome-Wide Association Study of the Genetic Determinants of Emphysema Distribution. *Am J Respir Crit Care Med*. 2017;195(6):757-771. doi:10.1164/rccm.201605-0997OC
41. Cho MH, Castaldi PJ, Wan ES, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet*. 2012;21(4):947-957. doi:10.1093/hmg/ddr524
42. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet*. 2019;51(3):494-505. doi:10.1038/s41588-018-0342-2
43. Hobbs BD, de Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
44. Wain L V, Shrine N, Artigas MS, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential

druggable targets. *Nat Genet*. 2017;49(3):416-425. doi:10.1038/ng.3787

45. Pillai SG, Ge D, Zhu G, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. McCarthy MI, ed. *PLoS Genet*. 2009;5(3):e1000421. doi:10.1371/journal.pgen.1000421
46. Cho MH, Boutaoui N, Klanderman BJ, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet*. 2010;42(3):200-202. doi:10.1038/ng.535
47. Silverman EK. Applying functional genomics to chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2018;15(4):S239-S242. doi:10.1513/AnnalsATS.201808-530MG
48. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019;51(3):481-493. doi:10.1038/s41588-018-0321-7
49. Silverman EK, Province MA, Campbell EJ, Pierce JA, Rao DC, Boerwinkle E. Family study of α1-antitrypsin deficiency: Effects of cigarette smoking, measured genotype, and their interaction on pulmonary function and biochemical traits. *Genet Epidemiol*. 1992;9(5):317-331. doi:10.1002/gepi.1370090504
50. Castaldi PJ, Demeo DL, Hersh CP, et al. Impact of non-linear smoking effects on the identification of gene-by-smoking interactions in COPD genetics studies. *Thorax*. 2011;66(10):903-909. doi:10.1136/thx.2010.146118
51. Aschard H, Lutz S, Maus B, et al. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Hum Genet*. 2012;131(10):1591-1613. doi:10.1007/s00439-012-1192-0
52. Hancock DB, Soler Artigas MM, Gharib SA, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet*. 2012;8(12):e1003098. doi:10.1371/journal.pgen.1003098
53. Park B, Koo S-M, An J, et al. Genome-wide assessment of gene-by-smoking interactions in COPD. *Sci Rep*. 2018;8(1):9319. doi:10.1038/s41598-018-27463-5
54. Ahmad S, Varga T V, Franks PW. Gene x environment interactions in obesity: the state of the evidence. *Hum Hered*. 2013;75(2-4):106-115. doi:10.1159/000351070
55. Qi Q, Chu AY, Kang JH, et al. Sugar-Sweetened Beverages and Genetic Risk of Obesity. *N Engl J Med*. 2012;367(15):1387-1396. doi:10.1056/NEJMoa1203039
56. Garcia-Closas M, Rothman N, Figueroa JD, et al. Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Cancer Res*.

2013;73(7):2211-2220. doi:10.1158/0008-5472.CAN-12-2388

57. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. Polygenic Risk Score for Coronary Heart Disease Modifies the Elevated Risk by Cigarette Smoking for Disease Incidence. *Circ Genomic Precis Med*. 2018;11(1):e001856. doi:10.1161/CIRCGEN.117.001856
58. Aschard H, Tobin MD, Hancock DB, et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int J Epidemiol*. 2017;46(3):894-904. doi:10.1093/ije/dyw318
59. Regan E a, Hokanson JE, Murphy JR, et al. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD J Chronic Obstr Pulm Dis*. 2011;7(1):32-43. doi:10.3109/15412550903499522
60. Cho MH, McDonald M-LN, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med*. 2014;2(3):214-225. doi:10.1016/S2213-2600(14)70002-5
61. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643
62. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
63. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med*. 2017;195(5):557-582. doi:10.1164/rccm.201701-0218PP

Chapter 2. Genome-wide association study of longitudinal change in quantitative emphysema

Chapter 2. Genome-wide association study of longitudinal change in quantitative emphysema

Introduction

Chronic obstructive pulmonary disease (COPD), defined by the presence of airflow obstruction on spirometry, is the third leading cause of death worldwide [1]. Despite the increasing burden of the disease, there is no medication that clearly ameliorates the progression of COPD. In addition to lung function, progression of this disease can also be assessed by worsening emphysema. Emphysema, characterized by the destruction of lung parenchyma, is a key component of COPD and associated with increased morbidity and mortality. The presence of emphysema is variable among subjects with similar degrees of airflow obstruction.

Advances in computed tomography (CT) imaging provide the opportunity to assess the extent and progression of emphysema quantitatively, and to study its related risk factors. The progression of emphysema (as determined by quantitative CT imaging) was examined in smokers [2], patients with alpha1-antitrypsin deficiency (AATD) [3,4] and smokers with COPD [5]. Longitudinal change in quantitative emphysema is associated with spirometric measures of lung function, severity of COPD and ongoing smoking, and as such has been proposed as a marker of response to therapy for COPD [6].

Although smoking is a strong risk factor for emphysema, there is evidence of genetic influences on emphysema. The estimated heritability of quantitative emphysema measured by CT scan was approximately 30% [7]. The longitudinal decline in lung function is reported to be moderately heritable (forced expiratory volume in the first second (FEV1): 0.05, forced vital capacity (FVC): 0.18 and FEV1/FVC: 0.13) [8]. Previous genome-wide association studies (GWAS) have

identified multiple genetic variants significantly associated with cross-sectional measures of quantitative emphysema based on CT imaging [9–13]. However, genetic determinants of longitudinal change in quantitative emphysema remain largely unknown. Identification of genetic factors for emphysema progression may identify different or more specific pathways than an analysis of overall lung function decline, expand our understanding of the genetics of COPD and contribute to the development of new drug therapies to slow the loss of lung density.

Our study aims to examine genetic variants for association with change in quantitative emphysema measured by CT imaging from two longitudinal cohort studies: COPDGene and Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). The specific objectives of the study are as follows: 1) to conduct GWAS to identify genetic variants associated with change in quantitative emphysema measured by CT imaging, and 2) to examine the association with change in quantitative emphysema for genetic loci previously identified as significantly associated with cross-sectional quantitative emphysema, COPD or lung function.

Methods

Study description

We included current and ex-smokers from the Genetic Epidemiology of COPD (COPDGene) study, a multicenter longitudinal cohort enrolling Non-Hispanic Whites (NHW) and African Americans (AA), and the Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE) study, a 3-year longitudinal study. Detailed descriptions, including genotyping quality control, genotyping imputation, and quantitative imaging, have been previously published [14,15].

COPDGene recruited subjects between ages of 45 and 80 years with a minimum of 10 pack-years smoking history at baseline and conducted volumetric inspiratory CT scans of the chest. Approximately 5 years later, subjects were asked to return to repeat the chest CT scan and detailed questionnaires. CT measures were available on 5,093 subjects COPDGene subjects at both visit 1 and visit 2. Among them, we excluded subjects who underwent a lung surgical procedure (including lung transplant and/or lung reduction surgery), changed their smoking behavior (either quitting smoking or resuming smoking between visits 1 and 2) and/or were never smoking control subjects.

ECLIPSE recruited COPD cases and controls aged 40 to 75 years with a minimum of 10 pack-years smoking history at baseline. Subjects included in this analysis were self-reported White ethnicity. Chest CT scan was conducted at baseline, after 1 year and after 3 years in the ECLIPSE study. CT measures were available on 1,871 subjects who had at least two visits including the baseline. Among them, we excluded subjects who changed their smoking behavior during the study.

CT measures

Volumetric inspiratory CT scans of the chest were acquired at maximal inspiration following standardized coaching and practiced breath-holding. Quantitative image analysis was performed to generate quantitative measures of emphysema using commercially available software Thirona LungQ (Thirona, Nijmegen, The Netherlands) for COPDGene, and Pulmonary Workstation 2.0 (VIDA Diagnostics, Coralville, IA, USA) for ECLIPSE. Percent low attenuation area (≤-950HU) (%LAA-950) was calculated as the percent of lung voxels with density less than -950 Hounsfield Units (HU). Adjusted lung density (ALD) was quantified as the lung density at

the 15th percentile of the HU distribution adjusted for the predicted total lung volume on inspiratory CT [16]. %LAA-950 and ALD were the main CT measures for the analysis and were used to calculate the annual change in emphysema.

Genotyping

In COPDGene, subjects were genotyped on the Illumina Human Omni Express array (San Diego, CA) and in ECLIPSE, the Illumina HumanHap 550 V3 array (San Diego, CA). Genotyping quality control (QC) was performed following previously described guidelines to remove low quality subjects and markers [14,17]. Unobserved genotypes were imputed using Michigan Imputation Server with the Haplotype Reference Consortium (HRC) panel [18].

Candidate region selection

To explore the effect of previously reported variants associated with cross-sectional quantitative emphysema, COPD, or spirometric measures of lung function on change in quantitative emphysema, we curated a list of associated single nucleotide polymorphisms (SNPs) from recent published GWASs. Seven loci for cross-sectional quantitative emphysema (*DLC1, SERPINA1, HHIP, CHRNA3, AGER, SNRPF, BICD1)*[9,11,19], 85 loci for COPD [20,21] and 279 loci for spirometric measures of lung function (FEV1, FVC, FEV1/FVC and Peak expiratory flow (PEF)) were included in the analysis [22]. We selected the most significant SNP for each genetic locus. SNPs for cross-sectional quantitative emphysema, COPD and lung function are listed in **Supplementary Table 1.1**. Given the small effects of individual genetic variants, we hypothesized that a polygenic risk score (PRS) comprised of the combined set of variants might be associated with cross-sectional quantitative emphysema and the annual change in emphysema. For the 279 SNPs previously associated with lung function mostly identified in

European ancestry, we generated a PRS by summing the number of risk alleles for each SNP with the weight of the FEV1/FVC effect size, as previously described [22], in our study population.

Statistical analysis

With two quantitative emphysema measures, %LAA-950 and ALD, we calculated the annual change as the difference in emphysema measure between the latest visit and baseline divided by the duration of follow-up in years. In COPDGene, with two time-points, the change in emphysema was generated by the difference between the second visit and baseline. In ECLIPSE where there were three time points available for CT measures, we calculated the difference between the latest visit and baseline for subjects with at least two measures.

We performed linear regression on each phenotype adjusted for age, sex, pack-years, smoking status (continued current/former smokers), change in scanner make, and principal components of genetic ancestry using EPACTS software (version 3.3, http://genome.sph.umich.edu/wiki/EPACTS/) stratified by race. To reduce the outlier effects and obtain a more normal distribution of outcome measures, the annual change in %LAA-950 and ALD were inverse-normal transformed. To combine results from COPDGene NHW, COPDGene AA, and ECLIPSE Whites, a fixed effect meta-analysis with an inverse variance weighting was conducted using the METAL software (http://csg.sph.umich.edu/abecasis/metal/) [23]. To examine effects of PRS on baseline emphysema and the annual change in emphysema, we performed linear regression adjusted for the same covariates as the main GWAS analysis.

Results

Basic characteristics

Basic characteristics of subjects from each study are shown in **Table 1.1**. The sample size included 3,030 NHW and 1,158 AA from COPDGene and 1,397 from ECLIPSE. Follow-up on average was approximately 5.5 years among COPDGene subjects and 2.6 years among ECLIPSE subjects. In both cohorts, the mean annual change in %LAA-950 was positive and mean annual change in ALD negative, consistent with an overall progression in emphysema, while the variance of both measures was quite large. The average annual change in %LAA was 0.02 (0.71) in NHW and 0.10 (0.54) in AA from COPDGene, and 0.59 (2.13) from ECLIPSE. The average annual change in ALD was -0.04 (2.05) in NHW and -0.22 (2.27) in AA from COPDGene and -1.27 (3.63) from ECLIPSE. At baseline, subjects with %LAA-950 > 5% were present at higher in ECLIPSE (79.6%) than in COPDGene (NHW (34.4%) and AA (16.2%)).

GWAS summary results

In the genome-wide association analysis, we considered a meta-analysis of all subjects as our primary analysis. No genetic variants reached genome-wide significance (P<5e-08). Eighteen loci for change in %LAA-950 and 17 loci for change in ALD reached a suggestive significance (P<1e-05) level (**Supplementary Table 1.2** and **Supplementary Figure 1.1**). For the association with %LAA-950, rs13164530 in the *WWC1* gene was most significantly associated (P=4.32e-07). For the association with ALD, rs7940672 near the *WEE1* gene was most significantly associated (P=2.74e-06).

From the meta-analysis of European ancestry subjects (3,030 COPDGene NHW and 1,397 ECLIPSE Whites), 18 loci for change in %LAA-950 and 20 loci for change in ALD reached a suggestive significance level (P<1e-05) (**Supplementary Table 1.3**). rs115047317 near the *NXPH2* gene achieved borderline genome-wide significance for change in %LAA-950 (P=9.31e-

08) and rs146580149 near the *LRIG2* loci was most significantly associated with change in ALD (P=1.04e-06). Among 1,158 AA from COPDGene, 44 loci for change in %LAA-950 and 27 loci for change in ALD yielded suggestive significance (P<1e-05). rs146932748 in the *CAB39L* gene and rs6733971 near the *ASB3* gene were most significantly associated with change in %LAA-950 (P=1.97e-07) and with change in ALD (P=1.46e-07), respectively.

Candidate SNP look-up

To examine candidate SNPs, we used the meta-analysis of European ancestries (3,030 COPDGene NHW and 1,397 ECLIPSE Whites). We curated 360 SNPs, in total, each previously reported to be associated with cross-sectional quantitative emphysema, COPD or spirometric measures of lung function. We used a Bonferroni corrected p-values (P=1.39e-04) as a measure of statistical significance.

None of the SNPs previously associated with cross-sectional quantitative emphysema were significantly associated with the annual change in emphysema. However, for SNPs previously reported to be associated with COPD and spirometric measures of lung function, the variant in the *DSP* gene, rs2076295, was associated with the annual change in %LAA-950 (risk allele=T, β (SE) =0.09 (0.02), P=3.79e-05) (**Table 1.2**). This variant was also associated, albeit at reduced significance, with the annual change in ALD (β (SE) = -0.06 (0.02), P=2.88e-03). In the original scale of outcome measures (without inverse normal transformation), for each risk allele the annual change in %LAA-950 increased by 0.06% (SE=0.02, P=7.74e-04) and the annual change in ALD decreased by 0.15 g/L (SE=0.05, P=2.06e-03) (**Supplementary Table 1.5**). In AA subjects from COPDGene, this variant was not significantly associated (%LAA-950; P=8.30e-01, ALD: P=9.45e-01). In the stratified analyses by COPD status (COPD case defined as Global Initiative for Obstructive Lung Disease (GOLD) criteria >=2) and presence of

emphysema defined as %LAA-950>5%, this association was still significant and the direction was consistent across COPD status and presence of emphysema (**Supplementary Tables 1.4** and **1.5** and **Supplementary Figures 1.2** and **1.3**). Other variants that approached significance (P<0.05) for change in either %LAA-950 or ALD traits are shown in **Table 1.2**.

Association of lung function polygenic risk score

The association between PRS based on spirometric measures of lung function and the annual change in emphysema was examined in European ancestry subjects (3,030 COPDGene NHW and 1,397 ECLIPSE Whites). Scatter plots between the weighted PRS and quantitative measures of emphysema for each study are shown in **Figure 1.1**. For all study populations, higher PRS correlated with higher levels of emphysema at baseline. The adjusted mean annual change in %LAA-950 in the highest decile was 0.02% (SE=0.06, P=0.66) and 0.45% (SE=0.25, P=0.07) higher than that seen in the lowest decile in COPDGene NHW and ECLIPSE Whites, respectively (data not shown). The adjusted mean annual change in ALD in the highest decile was 0.25 g/L (SE=0.16, P=0.12) and 0.49 g/L (SE=0.43, P=0.26) lower than that seen in the lowest decile in COPDGene NHW and ECLIPSE Whites, respectively (data not shown).

We observed a significant association of annual change in emphysema with weighted PRS in European ancestry subjects (**Table 1.3**). The weighted PRS was positively associated with the annual change in %LAA-950 (β (SE) =0.0025 (0.0012), P=4.03e-02) and showed a trend toward negative association with the annual change in ALD (β (SE) = -0.0021 (0.0012), P=7.31e-02). In the original scale of outcome measures (without inverse normal transformation), per one unit increase in the weighted PRS, %LAA-950 annually increased by 0.001% (SE=0.001, P=2.13e-01) and ALD annually decreased by 0.0055 g/L (SE=0.0027, P=4.16e-02) (**Supplementary Table 1.6**).

Discussion

This is the first GWAS investigating change in emphysema quantitatively measured by CT imaging. We conducted a GWAS to identify variants associated with annual change in quantitative emphysema measured by CT scans and examined effects of variants previously associated with cross-sectional quantitative emphysema, COPD and spirometric measures of lung function on the annual change in emphysema from two large cohorts, COPDGene and ECLIPSE. None of SNPs yielded genome-wide significance. However, in our candidate region analysis, we identified significant associations of a variant in *DSP* and a PRS based on spirometric measures of lung function with the annual change in emphysema in European ancestry.

Interestingly in our candidate region analysis, a variant in the *DSP* gene (rs2076295 T>G) was associated with annual change in %LAA-950, passing the Bonferroni corrected significance. The effect size was similar across studies of European ancestry. This finding is particularly interesting since rs2076295 is associated with COPD (risk allele=T, P=4.95e-08, OR (95% CI) =1.11(1.07-1.15)) [21] and a spirometric measure of lung function (FEV1/FVC ratio) (risk allele=T, P=6.95e-23, β (SE) = -0.02 (0.002)) [22], as well as idiopathic pulmonary fibrosis (risk allele=G, P=1.14E-16, OR (95% CI) =1.43(1.32-1.55)) [24]. The T allele of rs2076295 is associated with greater progression of emphysema, higher risk of COPD, and lower lung function, but has a protective effect on pulmonary fibrosis.

rs2076295 is intronic to the *DSP* gene. The *DSP* gene encodes desmoplakin, a major protein in desmosomes which are critical to cell-cell adhesion [25]. Desmosomes mechanically connect cells

and stabilize tissue architecture [26]. Desmosomes are essential in cell proliferation, differentiation, migration, morphogenesis, and apoptosis [26]. Mutations in *DSP* have been linked to several Mendelian syndromes involving palmoplantar keratoderma [27], left ventricular cardiomyopathy [27], familial arrhythmogenic right ventricular dysplasia [28], and lethal ancantholytic epidermolysis bullosa [29]. The *DSP* gene is highly expressed in the airway epithelia [30]. rs2076295 in the *DSP* gene is associated with differential gene expression of idiopathic pulmonary fibrosis in human lung [24,30]. This region was not previously reported to be associated with quantitative emphysema in cross-sectional studies. Thus, whether this association represents true progression of emphysema or (for the opposite allele) development of fibrosis, or both needs confirmation by further studies.

To jointly examine the effect of 279 SNPs of lung function on the annual change in quantitative emphysema, we applied a previously constructed PRS weighted by FEV1/FVC effect sizes in European ancestry [22]. The PRS of lung function was strongly associated with baseline emphysema. These findings are consistent with the strong correlation between emphysema and lung function. In addition, we found suggestive evidence that individuals with higher PRS showed more rapid emphysema progression. Subjects in the highest decile of PRS showed a trend toward greater emphysema progression than those in the lowest decile, even though it was not statistically significant ($P>0.05$). Our findings suggest that genetic variants associated with cross-sectional lung function might affect development and progression of emphysema, and to our knowledge, are the first description of an association of a genetic risk score with disease progression in COPD.

Our study showed differences in spirometry grade and emphysema measures between two studies, COPDGene and ECLIPSE (**Table 1.1**). It may be due to the different imaging protocols; in COPDGene, volumetric inspiratory CT acquisitions were obtained at 200mAs [31], and in ECLIPSE [32] all subjects underwent a low-dose volumetric inspiratory CT scan at 40mAs. In addition, ECLIPSE included a higher percentage of COPD subjects than COPDGene.

Despite our finding in our candidate region analysis, we did not find significant variants in the genome-wide analysis. The failure to identify novel genome-wide significant variants is likely due to several factors. First, CT measures to quantify the extent of the emphysema inherently have large variation [33]. While previous successful GWASs for cross-sectional quantitative emphysema used the same CT measures as our current study [9,10], and previous epidemiologic studies have found significant longitudinal associations of ALD with severity of COPD [2], our sample size was smaller than these prior studies, and likely effect sizes for longitudinal change in emphysema are considerably smaller than for cross-sectional genetic association or epidemiologic analyses. In a genome-wide setting, more accurate CT emphysema measures may be needed to detect the genetic variants with relatively small effect size associated with longitudinal change in emphysema. Second, the follow-up period in our study may be too short to capture the natural history of emphysema in adults. Our follow-up period (approximately 5.5 years in COPDGene and 2.6 years in ECLIPSE) is sufficient to show emphysema progression, showing an increase in %LAA-950 and decrease in ALD, but very short in the context of the natural history of emphysema. To estimate the true emphysema trajectory in adults, such a follow-up period may not be sufficient. The emphysema progression in our population may have an episodic rather than gradual pattern of progression, which could interfere with our ability to detect genetic associations. Also, the observed annual change in emphysema may be influenced by the regression to the mean due to the short follow-up period and the

measurement error of CT measures. Longer follow-up may provide more accurate data to estimate the emphysema progression. Third, our study populations were enriched for COPD patients. Though we observed emphysema progression in our population, to measure the true emphysema trajectory, normal subjects without lung abnormalities at baseline may be required to detect the genetic variants associated with emphysema progression. Fourth, emphysema progression may be less heritable than cross-sectional quantitative measures of emphysema. Traits of progression (as demonstrated by the heritability of cross-sectional lung function, versus progression [8]) are likely to be less heritable than ones of development.

Genetic determinants of emphysema progression remain poorly understood. This is the first GWAS investigating change in quantitative emphysema measured by CT imaging. No genetic variants were associated with annual change in emphysema at genome-wide significance. Further study with larger sample sizes and longer follow-up periods may be required to identify genetic determinants of emphysema progression. However, we observed a significant association of a *DSP* variant*,* previously reported in idiopathic pulmonary fibrosis, COPD and spirometric measures of lung function, with change in quantitative emphysema over time. This finding represents the first genetic association with emphysema progression measured by CT scan. PRS based on spirometric measures of lung function may also predict emphysema progression. Additional investigation of the *DSP* gene, and improved genetic risk scores are likely to provide further insights into the disease progression in emphysema and COPD.

References

1. World Health Organization. Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000–2016. https://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html.

Published 2018.

2. Diaz AA, Strand M, Coxson HO, et al. Disease Severity Dependence of the Longitudinal Association Between CT Lung Density and Lung Function in Smokers. *Chest*. 2017;153(3):638-645. doi:10.1016/j.chest.2017.10.012
3. Dirksen A, Piitulainen E, Parr DG, et al. Exploring the role of CT densitometry: a randomised study of augmentation therapy in 1-antitrypsin deficiency. *Eur Respir J*. 2009;33(6):1345-1353. doi:10.1183/09031936.00159408
4. Stolk J, Stockley RA, Piitulainen E, Stoel BC. Relationship between Change in Lung Density and Long-Term Progression of Lung Function. *Am J Respir Crit Care Med*. 2015;192(1):114-116. doi:10.1164/rccm.201502-0370LE
5. Coxson HO, Dirksen A, Edwards LD, et al. The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: A prospective analysis from the ECLIPSE study. *Lancet Respir Med*. 2013;1(2):129-136. doi:10.1016/S2213-2600(13)70006-7
6. Chapman KR, Burdon JGW, Piitulainen E, et al. Intravenous augmentation treatment and lung density in severe α1 antitrypsin deficiency (RAPID): a randomised, double-blind, placebo-controlled trial. *Lancet (London, England)*. 2015;386(9991):360-368. doi:10.1016/S0140-6736(15)60860-1
7. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med*. 2013;188(8):941-947. doi:10.1164/rccm.201302-0263OC
8. Gottlieb DJ, Wilk JB, Harmon M, et al. Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med*. 2001;164(9):1655-1659. doi:10.1164/ajrccm.164.9.2010122
9. Manichaikul A, Hoffman EA, Smolonska J, et al. Genome-wide study of percent emphysema on computed tomography in the general population. The Multi-Ethnic Study of Atherosclerosis Lung/SNP Health Association Resource Study. *Am J Respir Crit Care Med*. 2014;189(4):408-418. doi:10.1164/rccm.201306-1061OC
10. Cho MH, Castaldi PJ, Hersh CP, et al. A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. *Am J Respir Crit Care Med*. 2015;192(5):559-569. doi:10.1164/rccm.201501-0148OC
11. Kong X, Cho MH, Anderson W, et al. Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. *Am J Respir Crit Care Med*. 2011;183(1):43-49.

doi:10.1164/rccm.201004-0541OC

12. Castaldi PJ, Cho MH, San José Estépar R, et al. Genome-wide association identifies regulatory Loci associated with distinct local histogram emphysema patterns. *Am J Respir Crit Care Med*. 2014;190(4):399-409. doi:10.1164/rccm.201403-0569OC
13. Boueiz A, Chang Y, Cho MH, et al. Lobar Emphysema Distribution Is Associated With 5-Year Radiological Disease Progression. *Chest*. 2018;153(1):65-76. doi:10.1016/j.chest.2017.09.022
14. Cho MH, Boutaoui N, Klanderman BJ, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet*. 2010;42(3):200-202. doi:10.1038/ng.535
15. Cho MH, Mcdonald M-LN, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome- wide association study and meta-analysis. *Lancet Respir Med*. 2014;21016(314):214-225. doi:10.1016/S2213-2600(14)70002-5
16. Stoel BC, Putter H, Bakker ME, et al. Volume Correction in Computed Tomography Densitometry for Follow-up Studies on Pulmonary Emphysema. *Proc Am Thorac Soc*. 2008;5(9):919-924. doi:10.1513/pats.200804-040QC
17. Cho MH, McDonald M-LN, Zhou X, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med*. 2014;2(3):214-225. doi:10.1016/S2213-2600(14)70002-5
18. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643
19. Cho MH, Castaldi PJ, Wan ES, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet*. 2012;21(4):947-957. doi:10.1093/hmg/ddr524
20. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet*. 2019;51(3):494-505. doi:10.1038/s41588-018-0342-2
21. Hobbs BD, de Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
22. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019;51(3):481-493. doi:10.1038/s41588-018-0321-7

23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340
24. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet*. 2013;45(6):613-620. doi:10.1038/ng.2609
25. Vasioukhin V, Bowers E, Bauer C, Degenstein L, Fuchs E. Desmoplakin is essential in epidermal sheet formation. *Nat Cell Biol*. 2001;3(12):1076-1085. doi:10.1038/ncb1201-1076
26. Huber O, Petersen I. 150th anniversary series: Desmosomes and the hallmarks of cancer. *Cell Commun Adhes*. 2015. doi:10.3109/15419061.2015.1039642
27. Norgett EE, Hatsell SJ, Carvajal-Huerta L, et al. Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma. *Hum Mol Genet*. 2000;9(18):2761-2766. doi:10.1093/hmg/9.18.2761
28. den Haan AD, Tan BY, Zikusoka MN, et al. Comprehensive desmosome mutation analysis in north americans with arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circ Cardiovasc Genet*. 2009;2(5):428-435. doi:10.1161/CIRCGENETICS.109.858217
29. Jonkman MF, Pasmooij AMG, Pasmans SGMA, et al. Loss of desmoplakin tail causes lethal acantholytic epidermolysis bullosa. *Am J Hum Genet*. 2005;77(4):653-660. doi:10.1086/496901
30. Mathai SK, Pedersen BS, Smith K, et al. Desmoplakin Variants Are Associated with Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2016;193(10):1151-1160. doi:10.1164/rccm.201509-1863OC
31. Regan E a, Hokanson JE, Murphy JR, et al. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD J Chronic Obstr Pulm Dis*. 2011;7(1):32-43. doi:10.3109/15412550903499522
32. Vestbo J, Anderson W, Coxson HO, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J*. 2008;31(4):869-873. doi:10.1183/09031936.00111707
33. Lynch DA, Al-Qaisi MA. Quantitative Computed Tomography in Chronic Obstructive Pulmonary Disease. *J Thorac Imaging*. 2013;28(5):284-290.

doi:10.1097/RTI.0b013e318298733c

Table 1.1. Subject Characteristics

	COPDGene		ECLIPSE
	NHW	AA	White
N	3030	1158	1397
Age (year)	62.42 (8.35)	54.45 (7.09)	62.78 (7.50)
Female	1500 (49.5)	579 (50.0)	489 (35.0)
Pack-years	44.54 (24.28)	37.19 (20.75)	47.70 (27.21)
Former smokers	2161 (71.3)	242 (20.9)	896 (64.1)
BMI (kg/m2)	29.08 (5.84)	29.30 (6.69)	26.64 (5.39)
Spirometry Grade			
PRISm	317 (10.5)	177 (15.4)	NA
Control	1363 (45.1)	645 (56.3)	146 (10.5)
GOLD I	314 (10.4)	69 (6.0)	NA
GOLD II	641 (21.2)	175 (15.3)	533 (38.2)
GOLD III	312 (10.3)	69 (6.0)	551 (39.4)
GOLD IV	77 (2.5)	11 (1.0)	167 (12.0)
Annual change in percent emphysema	0.02 (0.71)	0.10 (0.54)	0.59 (2.13)
Annual change in adjusted lung density	-0.04 (2.05)	-0.22 (2.27)	-1.27 (3.63)
Follow-up Period (years)	5.51 (0.71)	5.59 (0.89)	2.56 (0.81)
Emphysema case (%LAA-950 > 5%) at baseline	1041 (34.4)	188 (16.2)	1112 (79.6)

Mean (SD) for continuous variable; N (%) for categorical variable; GOLD=Global Initiative for Chronic Obstructive Lung Disease; PRISm=Preserved Ratio Impaired Spirometry (FEV1 < 80% predicted with FEV1/FVC > 0.7); NHW=Non-Hispanic Whites; AA=African Americans; ECLIPSE= Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points

Table 1.2. Association of previously associated variants with annual change in emphysema in candidate gene analysis

							Meta-analysis of Whites			African Americans		
Previously associated trait	**Nearest gene**	**Chr.**	**Position**	**SNP**	**Risk allele**	**Alt. allele**	**Beta**	**SE**	**P**	**Beta**	**SE**	**P**
%LAA-950												
COPD,Lung function	*DSP*	6	7563232	rs2076295	T	G	0.086	0.021	3.79E-05	0.008	0.038	8.30E-01
Lung function	*LY86*	6	6741932	rs1294417	T	C	0.062	0.021	2.83E-03	-0.002	0.041	9.69E-01
Lung function	*TGFB2*	1	218631452	rs6604614	C	G	0.065	0.022	3.82E-03	0.042	0.039	2.85E-01
Lung function	*BMP4*	14	54419106	rs35107139	C	A	0.063	0.022	4.17E-03	-0.056	0.04	1.56E-01
Lung function	*GLIS2*	16	4361138	rs56104880	C	T	0.064	0.022	4.67E-03	-0.009	0.046	8.52E-01
COPD	*VGLL4*	3	11640601	rs2442776	G	A	0.083	0.029	4.90E-03	0.075	0.046	1.04E-01
Lung function	*VAPA*	18	10078071	rs8089099	A	G	0.065	0.023	5.03E-03	0.031	0.058	5.90E-01
COPD	*DENND2D*	1	111738108	rs629619	T	C	0.07	0.026	7.57E-03	-0.06	0.049	2.20E-01
Lung function	*CSF2*	5	131466629	rs3843503	T	A	0.053	0.021	1.17E-02	0.074	0.052	1.57E-01
Lung function	*JCAD*	10	30268770	rs7914842	G	A	0.05	0.021	1.68E-02	0.11	0.047	1.92E-02
Lung function	*CCDC91*	12	28588242	rs7977418	C	T	0.049	0.02	1.80E-02	-0.007	0.041	8.55E-01
Lung function	*AFAP1*	4	7879027	rs62289340	C	T	0.049	0.021	1.82E-02	-0.006	0.049	9.00E-01
COPD	*SERP2*	13	44842503	rs9525927	A	G	0.059	0.026	2.40E-02	0.002	0.052	9.71E-01
Lung function	*RPAP1*	15	41840238	rs2012453	G	A	0.045	0.02	2.77E-02	0.017	0.039	6.57E-01
COPD	*SPPL2C*	17	43924200	rs12373142	C	G	0.055	0.025	3.17E-02	0.026	0.099	7.94E-01
Lung function	*SUCLG2*	3	67455803	rs4132748	C	T	0.048	0.023	3.44E-02	0.005	0.05	9.12E-01
Lung function	*RIN3*	14	93098339	rs11621587	G	C	0.055	0.027	3.88E-02	-0.007	0.105	9.45E-01
Lung function	*MET*	7	116431427	rs193686	T	C	0.045	0.022	4.24E-02	0.008	0.038	8.37E-01
Lung function	*SPPL2C*	17	43940021	rs79412431	G	A	0.052	0.026	4.30E-02	0.054	0.098	5.82E-01
COPD	*RIN3*	14	93105953	rs72699855	G	C	0.053	0.026	4.41E-02	0.018	0.064	7.76E-01
Lung function	*LTBP4*	19	41117300	rs34093919	G	A	0.193	0.096	4.56E-02	NA	NA	NA

Lung function	*THSD4*	15	71803450	rs62015883	T	C	0.054	0.027	4.68E-02	-0.003	0.052	9.55E-01
COPD	*MMP3*	11	102720945	rs626750	G	A	0.053	0.027	4.72E-02	0.016	0.043	7.05E-01
ALD												
COPD,Lung function	*DSP*	6	7563232	rs2076295	T	G	-0.061	0.02	2.88E-03	-0.003	0.037	9.45E-01
Lung function	*GLIS2*	16	4361138	rs56104880	C	T	-0.059	0.022	7.50E-03	-0.06	0.044	1.72E-01
Lung function	*THSD4*	15	71803450	rs62015883	T	C	-0.07	0.027	9.01E-03	-0.09	0.05	7.06E-02
Emphysema	*SNRPF*	12	96260474	rs7957346	C	A	-0.046	0.02	2.42E-02	-0.027	0.038	4.87E-01
Lung function	*CSF2*	5	131466629	rs3843503	T	A	-0.046	0.021	2.46E-02	-0.038	0.05	4.45E-01
Lung function	*KIAA2012*	2	202970250	rs12997625	C	T	-0.045	0.02	2.58E-02	0.058	0.043	1.76E-01
COPD	*MECOM*	3	168746145	rs7642001	G	A	-0.046	0.021	2.75E-02	0.044	0.042	2.97E-01
Lung function	*TGFB2*	1	218855029	rs28613267	C	G	-0.044	0.02	2.82E-02	-0.003	0.038	9.33E-01
Lung function	*JCAD*	10	30268770	rs7914842	G	A	-0.045	0.021	2.95E-02	-0.021	0.046	6.48E-01
Lung function	*ATAD2B*	2	24018480	rs13009582	G	A	-0.044	0.02	3.05E-02	-0.009	0.041	8.31E-01
Lung function	*RPAP1*	15	41840238	rs2012453	G	A	-0.043	0.02	3.17E-02	-0.019	0.038	6.23E-01
Lung function	*LTBP4*	19	41117300	rs34093919	G	A	-0.2	0.095	3.49E-02	NA	NA	NA
Lung function	*IGFBP3*	7	46448518	rs17232687	C	T	-0.041	0.02	4.12E-02	0.064	0.047	1.73E-01
Lung function	*DEFB136*	8	11823332	rs4128298	C	T	-0.047	0.023	4.17E-02	0.05	0.047	2.83E-01
Lung function	*LY86*	6	6741932	rs1294417	T	C	-0.042	0.02	4.18E-02	0.025	0.04	5.32E-01
Lung function	*DHDDS*	1	26775367	rs9438626	G	C	-0.049	0.025	4.81E-02	-0.028	0.037	4.42E-01
Lung function	*SPAG17*	1	118911295	rs35043843	G	T	-0.048	0.024	4.81E-02	-0.068	0.051	1.88E-01
Emphysema	*SERPINA1*	14	94844947	rs28929474	T	C	-0.129	0.066	4.95E-02	NA	NA	NA

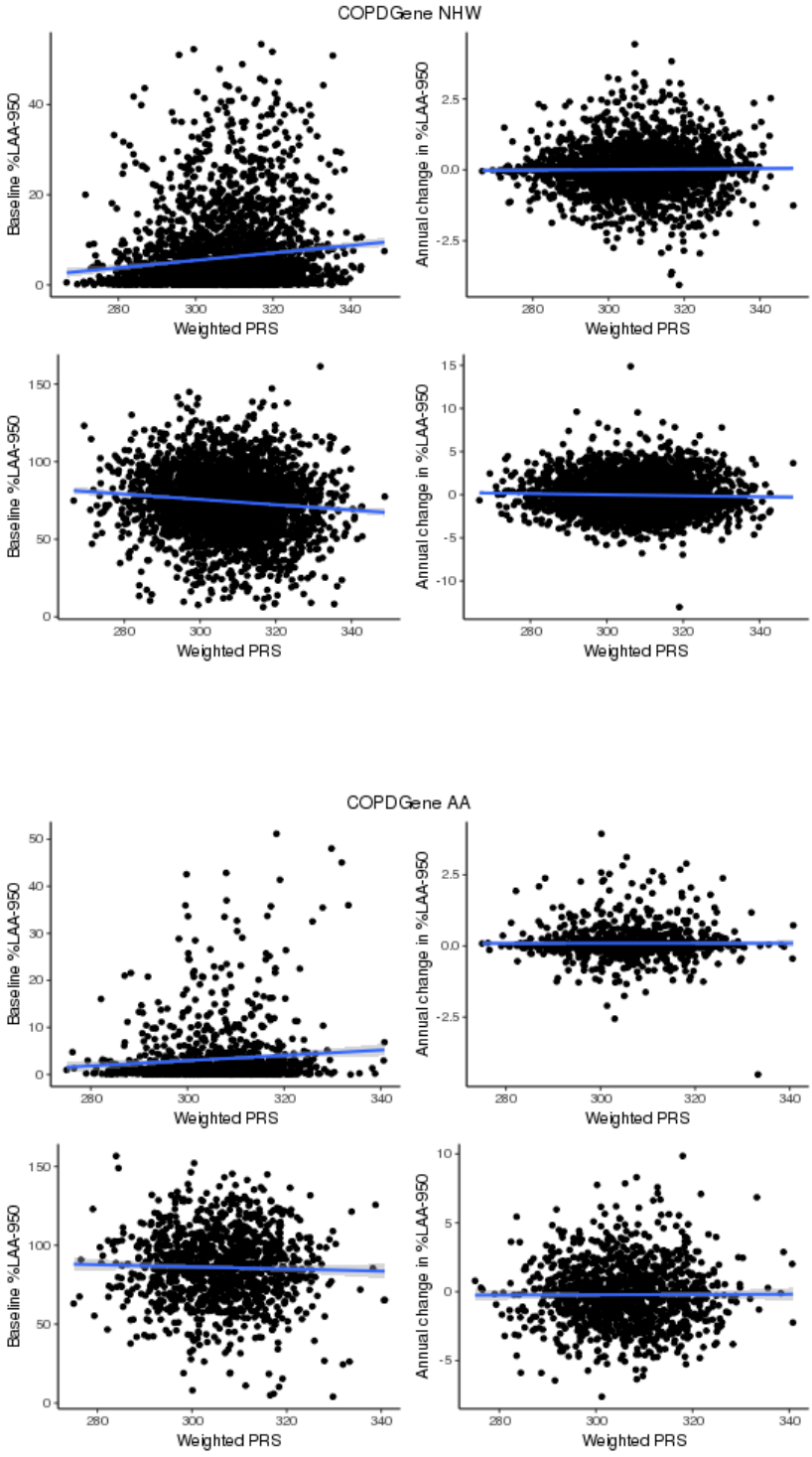
%LAA-950 = percentage of low-attenuation area less than -950 Hounsfield units; *ALD* = Adjusted lung density; *Chr.* = Chromosome; *Alt.allele* = Alternative allele

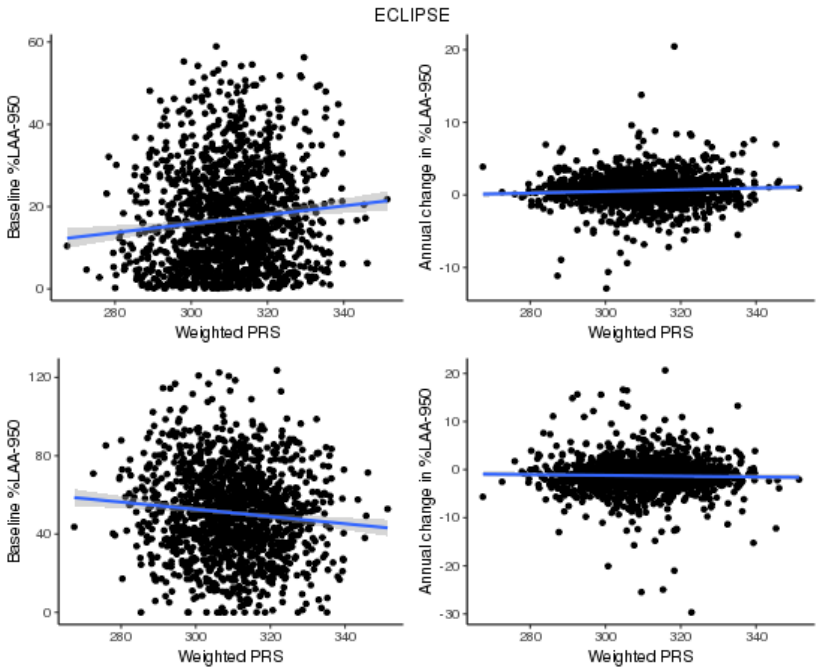
Table 1.3. Association of lung function polygenic risk score with baseline emphysema and annual change in emphysema

	Baseline emphysema			Annual change in emphysema		
	Beta	**SE**	**P**	**Beta**	**SE**	**P**
%LAA-950						
Meta-Analysis of Whites	0.0103	0.0011	3.00E-19	0.0025	0.0012	4.03E-02
COPDGene NHW	0.0109	0.0014	1.39E-15	0.0018	0.0014	2.10E-01
ECLIPSE White	0.0087	0.0021	4.63E-05	0.0041	0.0022	6.49E-02
COPDGene AA	0.0056	0.0026	2.98E-02	0.0013	0.0028	6.48E-01
ALD						
Meta-Analysis of Whites	-0.0075	0.0011	2.42E-11	-0.0021	0.0012	7.31E-02
COPDGene NHW	-0.0077	0.0013	8.06E-09	-0.0028	0.0014	4.37E-02
ECLIPSE White	-0.0071	0.0021	8.33E-04	-0.0003	0.0022	8.76E-01
COPDGene AA	-0.0039	0.0024	1.06E-01	0.0016	0.0027	5.54E-01

Outcome inversely normal transformed; %LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density; NHW=Non-Hispanic Whites; AA=African Americans; ECLIPSE= Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points

Figure 1.1. Scatter plot between lung function polygenic risk score and emphysema measures





%LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density; NHW=Non-Hispanic Whites; AA=African Americans; ECLIPSE= Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points; PRS=Polygenic Risk Score

Supplementary Table 1.1. List of SNPs for candidate region analysis

Previously associated trait	Nearest gene	Chr.	Position	SNP	Risk allele	Alt. allele	Risk Allele Frequency		
							COPDGene NHW	ECLIPSE Whites	COPDGene AA
Lung function	*PHF13*	1	6678864	rs9661802	C	A	0.34	0.35	0.34
COPD	*MFAP2*	1	17306029	rs9435731	C	A	0.48	0.49	0.81
Lung function	*MFAP2*	1	17308254	rs9435733	T	C	0.48	0.49	0.81
Lung function	*WNT4*	1	22612690	rs12737805	A	G	0.79	0.77	0.82
Lung function	*DHDDS*	1	26775367	rs9438626	G	C	0.79	0.81	0.47
Lung function	*DHDDS*	1	26796922	rs12096239	C	G	0.27	0.27	0.13
Lung function	*BMP8A*	1	39995074	rs755249	T	C	0.24	0.24	0.05
COPD	*PABPC4*	1	40060025	rs76841360	A	G	0.23	0.24	0.08
COPD	*TESK2*	1	45946636	rs4660861	T	G	0.44	0.43	0.37
Lung function	*FAF1*	1	51243374	rs1416685	C	G	0.42	0.4	0.56
COPD	*C1orf87*	1	60913143	rs72673419	C	T	0.96	0.95	0.9
Lung function	*NFIA*	1	60966772	rs72673461	T	G	0.96	0.95	0.92
Lung function	*NEXN*	1	78387270	rs9661687	C	T	0.14	0.15	0.28
Lung function	*TGFBR3*	1	92077097	rs1192415	G	A	0.19	0.19	0.2
Lung function	*TGFBR3*	1	92106637	rs10874851	A	C	0.49	0.48	0.85
Lung function	*TGFBR3*	1	92381483	rs11165787	G	A	0.31	0.34	0.11
Lung function	*DENND2D*	1	111737398	rs9970286	G	A	0.69	0.7	0.7
COPD	*DENND2D*	1	111738108	rs629619	T	C	0.19	0.18	0.21
Lung function	*SPAG17*	1	118911295	rs35043843	G	T	0.23	0.22	0.15
Lung function	*C1orf54*	1	150249101	rs11205354	A	C	0.45	0.44	0.85
Lung function	*MCL1*	1	150547747	rs878471	G	A	0.44	0.42	0.69
Lung function	*KRTCAP2*	1	155137395	rs141942982	G	T	0.89	0.89	0.69

Lung function	*RALGPS2*	1	178719306	rs4651005	C	T	0.7	0.67	0.9
Lung function	*HMCN1*	1	186090370	rs2146098	A	G	0.64	0.64	0.28
Lung function	*HMCN1*	1	186113852	rs17531405	G	C	0.81	0.82	0.96
Lung function	*PTPRC*	1	198898157	rs10919604	G	A	0.41	0.41	0.38
Lung function	*NR5A2*	1	200069216	rs2816992	A	G	0.59	0.59	0.69
Lung function	*LMOD1*	1	201884647	rs4309038	C	G	0.45	0.44	0.25
Lung function	*PIK3C2B*	1	204426295	rs1008833	G	A	0.14	0.15	0.03
Lung function	*KCNK2*	1	215120596	rs556648	G	A	0.77	0.76	0.22
Lung function	*TGFB2*	1	218521609	rs2799098	A	G	0.82	0.83	0.94
Lung function	*TGFB2*	1	218631452	rs6604614	C	G	0.71	0.74	0.43
COPD	*TGFB2*	1	218689155	rs3009947	C	T	0.51	0.53	0.47
Lung function	*TGFB2*	1	218855029	rs28613267	C	G	0.5	0.51	0.46
Lung function	*LYPLAL1*	1	219483218	rs75128958	G	A	0.92	0.91	0.99
Lung function	*SLC30A10*	1	219853742	rs1338227	G	T	0.45	0.45	0.84
COPD	*SLC30A10*	1	219924894	rs11118406	A	T	0.73	0.73	0.4
Lung function	*HLX*	1	221204299	rs17009288	A	C	0.7	0.71	0.72
Lung function	*DUSP10*	1	221631938	rs12757436	A	G	0.33	0.33	0.06
Lung function	*CHRM3*	1	239857524	rs2355237	A	G	0.5	0.51	0.63
COPD	*CHRM3*	1	239901006	rs11579382	G	C	0.56	0.58	0.67
COPD	*ASAP2*	2	9290357	rs955277	T	C	0.62	0.61	0.32
COPD	*DDX1*	2	15906179	rs10929386	T	C	0.52	0.53	0.51
Lung function	*DDX1*	2	15906854	rs2544536	C	T	0.53	0.54	0.54
Lung function	*KCNS3*	2	18287623	rs55884799	C	T	0.18	0.16	0.36
Lung function	*KCNS3*	2	18570024	rs6751968	A	C	0.18	0.18	0.41
Lung function	*RDH14*	2	18702313	rs13430465	T	C	0.08	0.08	0.09
Lung function	*ATAD2B*	2	24018480	rs13009582	G	A	0.54	0.55	0.71

Lung function	*CIB4*	2	26842146	rs732990	G	C	0.56	0.55	0.33
Lung function	*PKDCC*	2	42243850	rs4952564	A	G	0.69	0.69	0.37
COPD	*EML4*	2	42433247	rs12466981	T	C	0.27	0.28	0.11
Lung function	*EFEMP1*	2	56096892	rs3791679	A	G	0.75	0.76	0.93
Lung function	*IL1RL1*	2	102926362	rs12470864	G	A	0.61	0.62	0.83
Lung function	*CCNT2*	2	135672187	rs62168891	C	T	0.53	0.54	0.3
Lung function	*ZEB2*	2	145797829	rs1406225	T	G	0.29	0.29	0.41
COPD	*NR4A2*	2	157013035	rs72902175	C	T	0.88	0.87	0.96
Lung function	*NR4A2*	2	157016257	rs72902177	C	T	0.88	0.87	0.93
Lung function	*RBMS1*	2	161276378	rs7424771	A	G	0.43	0.45	0.43
Lung function	*OSBPL6*	2	179260382	rs2304340	A	G	0.41	0.4	0.27
Lung function	*ITGAV*	2	187530520	rs2084448	C	T	0.3	0.31	0.22
Lung function	*PLCL1*	2	199723365	rs1249096	G	A	0.43	0.41	0.37
Lung function	*SPATS2L*	2	201208692	rs985256	C	A	0.78	0.79	0.28
Lung function	*KIAA2012*	2	202970250	rs12997625	T	C	0.51	0.52	0.25
Lung function	*IGFBP5*	2	217614730	rs6435952	A	T	0.15	0.14	0.29
Lung function	*TNS1*	2	218604356	rs4294980	G	A	0.21	0.21	0.57
COPD,Lung function	*TNS1*	2	218683154	rs2571445	G	A	0.61	0.6	0.81
Lung function	*ASIC4*	2	220382700	rs4674407	C	T	0.5	0.5	0.59
Lung function	*PID1*	2	229502197	rs62201738	C	A	0.07	0.06	0.06
COPD	*PID1*	2	229569919	rs16825267	C	G	0.93	0.94	0.93
Lung function	*ASB1*	2	239441308	rs6710301	C	A	0.86	0.86	0.74
Lung function	*TWIST2*	2	239604970	rs6431620	T	G	0.77	0.79	0.69
COPD	*TWIST2*	2	239872704	rs62191105	T	C	0.2	0.19	0.04
Lung function	*TWIST2*	2	239881309	rs4308141	G	C	0.2	0.19	0.04
Lung function	*C2orf54*	2	241844033	rs6437219	C	T	0.49	0.49	0.6

Lung function	*BOK*	2	242495953	rs6733504	A	G	0.54	0.55	0.46
COPD	*VGLL4*	3	11640601	rs2442776	G	A	0.14	0.14	0.23
Lung function	*WNT7A*	3	13787641	rs2974389	A	G	0.43	0.43	0.52
Lung function	*RARB*	3	25179533	rs73048404	T	G	0.87	0.86	0.98
COPD,Lung function	*RARB*	3	25520582	rs1529672	C	A	0.83	0.83	0.81
Lung function	*RBMS3*	3	29469675	rs17666332	T	G	0.71	0.69	0.93
COPD	*RBMS3*	3	29472412	rs13073544	G	C	0.72	0.69	0.81
Lung function	*CACNA2D3*	3	55152319	rs12715478	A	G	0.59	0.59	0.59
COPD	*CACNA2D3*	3	55158224	rs17759204	A	G	0.73	0.71	0.87
COPD	*SLMAP*	3	57746515	rs62259026	T	C	0.24	0.23	0.36
Lung function	*SLMAP*	3	57879611	rs6445932	G	T	0.25	0.24	0.44
Lung function	*SUCLG2*	3	67455803	rs4132748	C	T	0.71	0.71	0.81
Lung function	*FOXP1*	3	71583177	rs35480566	A	G	0.58	0.58	0.8
Lung function	*PDZRN3*	3	73862616	rs586936	A	G	0.41	0.4	0.44
Lung function	*DCBLD2*	3	98822050	rs12497779	T	G	0.24	0.24	0.29
Lung function	*COL8A1*	3	99420192	rs1610265	T	C	0.07	0.07	0.17
COPD	*ADCY5*	3	123077042	rs4093840	T	A	0.54	0.51	0.37
Lung function	*EEFSEC*	3	127931340	rs2999090	A	G	0.88	0.89	0.89
COPD	*EEFSEC*	3	127961178	rs2955083	A	T	0.88	0.89	0.89
COPD	*ZBTB38*	3	141147414	rs7650602	C	T	0.43	0.43	0.7
Lung function	*RSRC1*	3	158226886	rs12634907	G	A	0.33	0.33	0.16
Lung function	*BCHE*	3	165548529	rs1799807	C	T	0.02	0.01	NA
Lung function	*MECOM*	3	168709843	rs879394	G	T	0.77	0.76	0.78
COPD	*MECOM*	3	168746145	rs7642001	G	A	0.62	0.61	0.74
Lung function	*MECOM*	3	169295436	rs78101726	A	G	0.85	0.86	0.87
Lung function	*IGF2BP2*	3	185503456	rs6780171	A	T	0.32	0.31	0.56

Lung function	*AFAP1*	4	7879027	rs62289340	C	T	0.56	0.56	0.81
Lung function	*KDR*	4	56012149	rs12331869	A	G	0.17	0.18	0.37
COPD	*BTC*	4	75673363	rs4585380	G	A	0.74	0.76	0.75
Lung function	*BTC*	4	75676529	rs62316310	G	A	0.74	0.76	0.75
Lung function	*FRAS1*	4	79403952	rs11098196	T	G	0.52	0.53	0.25
Lung function	*FAM13A*	4	89855495	rs2609279	C	T	0.8	0.8	0.74
Lung function	*FAM13A*	4	89869078	rs2869966	T	C	0.4	0.44	0.59
COPD	*FAM13A*	4	89883818	rs7671261	A	G	0.54	0.57	0.68
Lung function	*TET2*	4	106133184	rs6533183	C	T	0.38	0.36	0.34
Lung function	*INTS12*	4	106766430	rs11722225	T	C	0.93	0.93	0.96
COPD,Lung function	*NPNT*	4	106819053	rs34712979	G	A	0.76	0.73	0.95
Lung function	*HHIP*	4	145330628	rs13109426	G	A	0.4	0.41	0.77
Lung function	*HHIP*	4	145442364	rs13116999	G	A	0.45	0.45	0.21
COPD	*HHIP*	4	145489098	rs13140176	A	G	0.6	0.6	0.9
Emphysema,Lung function	*HHIP*	4	145506456	rs13141641	T	C	0.59	0.59	0.89
Lung function	*HHIP*	4	145740898	rs2353940	T	C	0.76	0.75	0.95
Lung function	*CEP72*	5	609661	rs11739847	A	G	0.19	0.17	0.03
Lung function	*TARS*	5	33352738	rs268717	C	T	0.08	0.09	0.24
Lung function	*NNT*	5	43976162	rs4866846	A	G	0.17	0.14	0.41
Lung function	*FGF10*	5	44367221	rs6859730	A	T	0.33	0.31	0.72
Lung function	*ITGA1*	5	52187038	rs12522114	C	A	0.73	0.74	0.79
COPD	*ITGA1*	5	52195033	rs1551943	A	G	0.23	0.22	0.07
Lung function	*ARL15*	5	53444498	rs2441026	T	C	0.45	0.49	0.39
COPD	*TNPO1*	5	72144005	rs34651	C	T	0.08	0.08	0.03
Lung function	*AP3B1*	5	77396400	rs425102	T	G	0.76	0.77	0.72

Lung function	*SPATA9*	5	95025146	rs987068	C	G	0.69	0.71	0.93
COPD	*SPATA9*	5	95036700	rs153916	T	C	0.54	0.56	0.6
Lung function	*SRFBP1*	5	121410529	rs10059661	C	G	0.83	0.83	0.9
Lung function	*ADAMTS19*	5	128767384	rs17163397	G	A	0.12	0.13	0.12
Lung function	*CSF2*	5	131466629	rs3843503	T	A	0.57	0.57	0.82
COPD	*HSPA4*	5	132439010	rs62375246	A	T	0.26	0.27	0.31
COPD	*HTR4*	5	147854970	rs10037493	C	T	0.56	0.59	0.5
Lung function	*HTR4*	5	147856522	rs7733410	G	A	0.56	0.59	0.66
Lung function	*ADRB2*	5	148206885	rs1800888	T	C	0.01	0.01	NA
Lung function	*AFAP1L1*	5	148652302	rs11952673	G	T	0.62	0.61	0.82
COPD	*CCDC69*	5	150595073	rs979453	A	G	0.67	0.66	0.3
Lung function	*ADAM19*	5	156908317	rs11134766	C	T	0.94	0.93	0.99
COPD	*ADAM19*	5	156937043	rs10866659	G	A	0.34	0.35	0.59
Lung function	*ADAM19*	5	156944199	rs11134789	A	C	0.34	0.35	0.57
Lung function	*FGF18*	5	170901463	rs10059996	T	G	0.37	0.39	0.73
COPD	*FGF18*	5	170901586	rs12519165	T	A	0.61	0.6	0.27
Lung function	*RASGEF1C*	5	179598771	rs79898473	C	T	0.32	0.31	0.16
Lung function	*LY86*	6	6741932	rs1294417	T	C	0.45	0.47	0.65
COPD	*RREB1*	6	7211818	rs1334576	A	G	0.41	0.41	0.35
COPD,Lung function	*DSP*	6	7563232	rs2076295	T	G	0.55	0.56	0.5
Lung function	*BMP6*	6	7720059	rs12198986	A	G	0.46	0.45	0.23
Lung function	*BMP6*	6	7797840	rs10498672	C	G	0.83	0.83	0.94
COPD	*ID4*	6	19842661	rs9350191	C	T	0.15	0.14	0.13
COPD	*PRL*	6	22004909	rs13198656	C	T	0.46	0.45	0.82
Lung function	*PRL*	6	22017543	rs13198081	C	G	0.37	0.37	0.33
Lung function	*ZSCAN31*	6	28301099	rs7752448	A	G	0.89	0.9	0.71

COPD	*IER3*	6	30713580	rs2284174	C	T	0.21	0.22	0.48
Emphysema,COPD,Lung function	*AGER*	6	32151443	rs2070600	C	T	0.96	0.96	NA
Lung function	*HMGA1*	6	34188892	rs9689096	A	C	0.94	0.94	0.75
Lung function	*CDC5L*	6	44447598	rs9357446	A	G	0.53	0.52	0.83
Lung function	*RUNX2*	6	45530471	rs12202314	T	C	0.68	0.67	0.43
Lung function	*RUNX2*	6	45622748	rs9472541	T	A	0.7	0.71	0.19
Lung function	*DST*	6	56336406	rs2894837	G	A	0.37	0.36	0.63
Lung function	*KCNQ5*	6	73663814	rs13206405	A	C	0.19	0.2	0.13
COPD	*ARMC2*	6	109266255	rs2806356	T	C	0.81	0.8	0.95
Lung function	*ARMC2*	6	109268050	rs2798641	C	T	0.81	0.8	0.95
COPD	*RFX6*	6	117257018	rs674621	T	C	0.68	0.69	0.66
Lung function	*CENPW*	6	126990392	rs6918725	T	G	0.49	0.48	0.68
Lung function	*SLC2A12*	6	134339265	rs2627237	G	A	0.41	0.41	0.48
Lung function	*CITED2*	6	140271357	rs1102077	C	A	0.25	0.25	0.14
COPD	*CITED2*	6	140280398	rs646695	C	T	0.25	0.25	0.16
Lung function	*VTA1*	6	142560957	rs9385988	G	A	0.26	0.25	0.24
COPD	*ADGRG6*	6	142668901	rs9399401	C	T	0.27	0.26	0.39
Lung function	*ADGRG6*	6	142688969	rs17280293	A	G	0.97	0.98	NA
Lung function	*ADGRG6*	6	142745883	rs7753012	G	T	0.31	0.3	0.85
COPD	*AMZ1*	7	2752152	rs798565	G	A	0.72	0.73	0.9
Lung function	*C1GALT1*	7	7256490	rs4318980	A	G	0.42	0.43	0.67
Lung function	*AGMO*	7	15506007	rs4721442	G	T	0.17	0.18	0.13
Lung function	*MEOX2*	7	15872324	rs4721457	C	T	0.15	0.15	0.14
COPD	*ITGB8*	7	20418134	rs2040732	T	C	0.43	0.41	0.34
Lung function	*SKAP2*	7	26848830	rs559233	C	T	0.5	0.5	0.41

Lung function	*HOXA3*	7	27182329	rs62454414	T	G	0.87	0.86	0.97
Lung function	*JAZF1*	7	28200097	rs1513272	C	T	0.49	0.51	0.75
Lung function	*IGFBP3*	7	46448518	rs17232687	C	T	0.51	0.5	0.19
Lung function	*SEMA3D*	7	84569510	rs12707691	G	C	0.32	0.31	0.14
COPD	*ZKSCAN1*	7	99630342	rs2897075	C	T	0.62	0.63	0.86
Lung function	*MCM7*	7	99692993	rs2261360	G	T	0.75	0.78	0.9
Lung function	*MET*	7	116431427	rs193686	T	C	0.69	0.67	0.6
Lung function	*RNF32*	7	156127246	rs12698403	A	G	0.44	0.43	0.28
COPD	*MFHAS1*	8	8697658	rs9329170	G	C	0.15	0.14	0.26
Lung function	*PPP1R3B*	8	9018590	rs330939	T	G	0.61	0.6	0.74
Lung function	*DEFB136*	8	11823332	rs4128298	C	T	0.28	0.26	0.25
Emphysema	*DLC1*	8	13054869	rs75200691	G	T	0.11	0.12	0.08
Lung function	*SULF1*	8	70367248	rs7465401	T	C	0.72	0.7	0.72
Lung function	*HSF1*	8	145504343	rs7838717	T	C	0.37	0.38	0.11
Lung function	*SMARCA2*	9	1568941	rs771662	T	C	0.34	0.35	0.59
Lung function	*GLIS3*	9	4120648	rs1570203	A	G	0.53	0.54	0.61
COPD	*GLIS3*	9	4143749	rs10114763	A	T	0.58	0.59	0.62
Lung function	*SH3GL2*	9	18013733	rs7041139	T	C	0.32	0.33	0.5
Lung function	*ELAVL2*	9	23587027	rs1107677	T	C	0.48	0.48	0.44
COPD	*ELAVL2*	9	23588684	rs156394	C	T	0.46	0.46	0.5
COPD	*RASEF*	9	85126163	rs7866939	C	T	0.32	0.35	0.58
Lung function	*PTCH1*	9	98266855	rs28446321	T	A	0.91	0.89	0.82
Lung function	*ERCC6L2*	9	98878881	rs72743974	G	A	0.16	0.17	0.15
Lung function	*GALNT12*	9	101632854	rs57649467	G	A	0.62	0.6	0.6
COPD	*COL15A1*	9	101661650	rs10760580	A	G	0.28	0.29	0.1
Lung function	*ZNF462*	9	109483517	rs1491106	T	G	0.37	0.37	0.38

Lung function	*ASTN2*	9	119234058	rs10983184	C	T	0.36	0.37	0.63
COPD	*ASTN2*	9	119401650	rs803923	A	G	0.54	0.56	0.37
Lung function	*IER5L*	9	131943843	rs967497	G	A	0.7	0.7	0.32
Lung function	*QSOX2*	9	139100413	rs7024579	T	C	0.31	0.3	0.07
Lung function	*CARD9*	9	139259349	rs4073153	G	A	0.43	0.44	0.32
COPD	*CDC123*	10	12277992	rs7068966	C	T	0.48	0.48	0.78
Lung function	*CDC123*	10	12278021	rs7090277	T	A	0.48	0.48	0.78
Lung function	*JCAD*	10	30268770	rs7914842	G	A	0.44	0.44	0.76
Lung function	*PARD3*	10	34480582	rs1274475	G	A	0.62	0.61	0.9
Lung function	*JMJD1C*	10	64998971	rs7082066	G	A	0.81	0.8	0.47
Lung function	*MYPN*	10	69962954	rs10998018	A	G	0.47	0.49	0.28
Lung function	*CAMK2G*	10	75580014	rs7098573	A	G	0.74	0.72	0.54
Lung function	*CAMK2G*	10	75639578	rs60820984	T	C	0.2	0.2	0.13
Lung function	*ZNF503*	10	77119039	rs1259605	T	C	0.75	0.77	0.92
Lung function	*LRMDA*	10	78312002	rs2637254	G	A	0.47	0.48	0.73
COPD	*LRMDA*	10	78318879	rs2579762	A	C	0.51	0.51	0.82
COPD,Lung function	*SFTPD*	10	81706324	rs721917	G	A	0.43	0.44	0.41
Lung function	*STN1*	10	105639611	rs11191841	C	T	0.5	0.5	0.37
COPD	*STN1*	10	105656874	rs1570221	G	A	0.65	0.65	0.84
Lung function	*DMBT1*	10	124297637	rs4279944	C	T	0.85	0.84	0.8
COPD	*ARNTL*	11	13171236	rs4757118	C	T	0.45	0.44	0.83
Lung function	*SLC1A2*	11	35308988	rs10836366	T	C	0.76	0.73	0.73
Lung function	*HSD17B12*	11	43690717	rs17596617	T	C	0.32	0.32	0.37
Lung function	*PRDM11*	11	45244903	rs10838435	G	C	0.85	0.85	0.41
Lung function	*EML3*	11	62370155	rs71490394	G	A	0.63	0.65	0.87
Lung function	*ARHGEF17*	11	73036179	rs2027761	C	T	0.89	0.9	0.66

COPD	*PRSS23*	11	86444761	rs117261012	A	G	0.85	0.84	0.97
Lung function	*PRSS23*	11	86448839	rs11234768	T	C	0.85	0.84	0.97
COPD	*MMP3*	11	102720945	rs626750	G	A	0.82	0.84	0.73
Lung function	*RPUSD4*	11	126009500	rs541601	C	T	0.81	0.81	0.62
Lung function	*FKBP4*	12	2908330	rs56196860	C	A	0.97	0.97	NA
Lung function	*CCND2*	12	4243749	rs12811814	T	C	0.45	0.46	0.68
Lung function	*AEBP2*	12	19808912	rs10841302	G	C	0.46	0.43	0.22
COPD	*CCDC91*	12	28320536	rs11049386	T	A	0.73	0.73	0.94
Lung function	*CCDC91*	12	28588242	rs7977418	C	T	0.46	0.46	0.3
Emphysema	*BICD1*	12	32380501	rs10844154	C	A	0.56	0.56	0.43
Lung function	*SUOX*	12	56396768	rs1689510	C	G	0.33	0.34	0.12
Lung function	*LRP1*	12	57527283	rs11172113	C	T	0.4	0.42	0.43
Lung function	*RASSF3*	12	65075332	rs1244869	G	T	0.37	0.37	0.23
Lung function	*MSRB3*	12	65793153	rs12825748	C	G	0.32	0.32	0.15
Lung function	*HMGA2*	12	66409367	rs11176001	C	A	0.87	0.87	0.96
Lung function	*ALX1*	12	85719906	rs56390486	A	G	0.29	0.29	0.17
Lung function	*CRADD*	12	94194890	rs9788269	G	A	0.25	0.26	0.14
Lung function	*FGD6*	12	95554771	rs113745635	C	T	0.78	0.76	0.81
COPD	*SNRPF*	12	96237570	rs7307510	T	C	0.19	0.16	0.25
Lung function	*SNRPF*	12	96242109	rs7970544	T	G	0.19	0.16	0.06
Emphysema	*SNRPF*	12	96260474	rs7957346	C	A	0.43	0.4	0.37
Lung function	*IGF1*	12	102824921	rs972936	T	C	0.27	0.26	0.37
Lung function	*TBX5*	12	114669870	rs2701110	C	A	0.85	0.84	0.88
Lung function	*TBX3*	12	115201436	rs10850377	A	G	0.34	0.33	0.17
Lung function	*TBX3*	12	115501127	rs35505	A	G	0.69	0.68	0.87
COPD	*MED13L*	12	115947901	rs7958945	G	A	0.36	0.37	0.69

Lung function	*SMIM2*	13	44820608	rs9533803	C	T	0.79	0.78	0.92
COPD	*SERP2*	13	44842503	rs9525927	A	G	0.81	0.82	0.83
Lung function	*KCNRG*	13	50707087	rs2812208	C	G	0.02	0.02	NA
Lung function	*KLHL1*	13	71647588	rs803765	A	C	0.35	0.35	0.17
Lung function	*NDFIP2*	13	80467235	rs4885681	C	T	0.28	0.27	0.7
Lung function	*DOCK9*	13	99665512	rs11620380	A	C	0.11	0.11	0.03
Lung function	*MYO16*	13	109918493	rs9634470	T	C	0.75	0.76	0.81
Lung function	*HAUS4*	14	23429729	rs1951121	G	T	0.4	0.42	0.48
Lung function	*BMP4*	14	54346010	rs74053129	A	G	0.1	0.09	0.03
Lung function	*BMP4*	14	54419106	rs35107139	C	A	0.42	0.4	0.5
Lung function	*VRTN*	14	74817418	rs10141786	A	G	0.42	0.42	0.19
Lung function	*FLRT2*	14	84338431	rs1756281	A	G	0.7	0.7	0.38
Lung function	*TRIP11*	14	92512143	rs11160037	A	G	0.61	0.63	0.34
Lung function	*RIN3*	14	93098339	rs11621587	G	C	0.81	0.83	0.96
COPD	*RIN3*	14	93105953	rs72699855	G	C	0.81	0.83	0.9
Emphysema	*SERPINA1*	14	94844947	rs28929474	T	C	0.02	0.03	NA
Lung function	*BMF*	15	40397191	rs34245505	C	G	0.81	0.79	0.96
Lung function	*IVD*	15	40716253	rs2304645	G	C	0.49	0.48	0.78
Lung function	*CHAC1*	15	41255396	rs4924525	A	C	0.51	0.52	0.36
Lung function	*RPAP1*	15	41840238	rs2012453	G	A	0.57	0.57	0.49
Lung function	*MGA*	15	41953211	rs56383987	T	C	0.05	0.06	NA
Lung function	*COPS2*	15	49409527	rs79234094	A	G	0.27	0.26	0.08
Lung function	*FAM227B*	15	49706145	rs35251997	T	A	0.07	0.07	0.05
COPD	*DTWD1*	15	49984710	rs72731149	C	G	0.07	0.07	0.24
Lung function	*USP3*	15	63866877	rs62012772	C	T	0.18	0.19	0.03
Lung function	*AAGAB*	15	67491274	rs12917612	A	C	0.23	0.22	0.08

COPD,Lung function	*THSD4*	15	71612514	rs1441358	T	G	0.68	0.65	0.49
Lung function	*THSD4*	15	71803450	rs62015883	T	C	0.18	0.17	0.17
Lung function	*REC114*	15	73833600	rs7176074	T	G	0.05	0.05	0.36
Emphysema,COPD	*CHRNA3*	15	78898932	rs55676755	G	C	0.37	0.42	0.17
Lung function	*SH3GL3*	15	84274591	rs1896797	A	G	0.49	0.48	0.84
COPD	*ADAMTSL3*	15	84392907	rs10152300	G	A	0.23	0.23	0.07
Lung function	*CLUAP1*	16	3583173	rs3751837	C	T	0.79	0.79	0.75
Lung function	*GLIS2*	16	4361138	rs56104880	C	T	0.31	0.29	0.68
Lung function	*GRIN2A*	16	10136889	rs11074547	T	G	0.74	0.72	0.61
COPD	*TEKT5*	16	10709013	rs56134392	C	T	0.34	0.35	0.63
Lung function	*TEKT5*	16	10740982	rs78442819	C	G	0.2	0.22	0.04
Lung function	*SH2B1*	16	28870962	rs12446589	A	G	0.39	0.4	0.09
Lung function	*TENT4B*	16	50188929	rs76219171	A	G	0.06	0.05	0.02
Lung function	*FTO*	16	53935407	rs35420030	T	C	0.95	0.95	0.99
COPD	*TEPP*	16	58022625	rs8044657	A	G	0.09	0.08	0.3
Lung function	*MMP15*	16	58063513	rs11648508	G	T	0.3	0.3	0.54
Lung function	*WWP2*	16	69891510	rs8047194	G	T	0.5	0.52	0.48
COPD	*CFDP1*	16	75340231	rs4888379	A	T	0.41	0.4	0.7
Lung function	*CFDP1*	16	75411445	rs11858992	A	C	0.41	0.4	0.64
Lung function	*WWOX*	16	78225633	rs2345443	A	G	0.31	0.31	0.17
Lung function	*FOXF1*	16	86403821	rs12918140	G	C	0.9	0.89	0.97
Lung function	*MTHFSD*	16	86579223	rs6539952	A	C	0.25	0.24	0.61
Lung function	*ATP2A3*	17	3882613	rs8082036	C	G	0.51	0.51	0.79
Lung function	*PITPNM3*	17	6469793	rs4796334	A	G	0.49	0.51	0.31
Lung function	*CLDN7*	17	7163350	rs1215	A	G	0.85	0.85	0.96
Lung function	*TNFSF12*	17	7448457	rs4968200	G	C	0.86	0.86	0.44

Lung function	*NCOR1*	17	16030520	rs34351630	T	C	0.45	0.46	0.7
Lung function	*SSH2*	17	28072327	rs2244592	A	G	0.47	0.47	0.2
COPD	*EFCAB5*	17	28413129	rs8080772	C	T	0.35	0.35	0.09
Lung function	*ATAD5*	17	29210595	rs62070648	A	G	0.27	0.27	0.1
COPD	*RPL23*	17	36835079	rs34727469	T	C	0.13	0.14	0.04
Lung function	*RPL23*	17	36915540	rs35246838	C	T	0.13	0.13	0.03
Lung function	*FBXL20*	17	37504933	rs8069451	C	T	0.27	0.25	0.7
COPD	*THRA*	17	38218773	rs62065216	A	G	0.42	0.45	0.26
COPD	*SPPL2C*	17	43924200	rs12373142	C	G	0.79	0.78	0.96
Lung function	*SPPL2C*	17	43940021	rs79412431	G	A	0.79	0.78	0.95
Lung function	*SKAP1*	17	46552229	rs12945803	C	T	0.22	0.22	0.14
Lung function	*ANKFN1*	17	54195453	rs28519449	T	C	0.39	0.4	0.2
Lung function	*BCAS3*	17	59286644	rs8068952	G	C	0.22	0.22	0.83
Lung function	*DDX5*	17	62497964	rs77672322	T	C	0.02	0.03	NA
Lung function	*SMURF2*	17	62686730	rs11653958	A	G	0.74	0.74	0.93
Lung function	*KCNJ2*	17	68976415	rs6501431	T	C	0.78	0.76	0.83
Lung function	*SOX9*	17	69201811	rs6501455	G	A	0.47	0.47	0.27
COPD	*SOX9*	17	69216687	rs11655567	T	C	0.53	0.53	0.71
Lung function	*SOX9*	17	69371318	rs996865	C	T	0.92	0.91	0.78
Lung function	*LLGL2*	17	73525670	rs9892893	T	G	0.26	0.27	0.33
Lung function	*ASPSCR1*	17	79952944	rs59606152	C	T	0.89	0.88	0.95
Lung function	*MTCL1*	18	8801351	rs513953	G	A	0.74	0.75	0.48
COPD	*MTCL1*	18	8808464	rs647097	T	C	0.72	0.73	0.61
Lung function	*VAPA*	18	10078071	rs8089099	A	G	0.27	0.27	0.13
Lung function	*GATA6*	18	19816712	rs1985511	A	T	0.46	0.45	0.43
Lung function	*RBBP8*	18	20234336	rs11082051	G	A	0.49	0.49	0.86

Lung function	*CABLES1*	18	20708321	rs9947743	A	G	0.79	0.79	0.93
Lung function	*RIOK3*	18	21074255	rs303752	A	G	0.4	0.4	0.11
Lung function	*HRH4*	18	22290711	rs1668091	T	C	0.69	0.68	0.64
Lung function	*SLC14A2*	18	42827898	rs9807668	C	T	0.91	0.9	0.97
Lung function	*DCC*	18	51022606	rs12607758	C	T	0.4	0.4	0.67
Lung function	*TCF4*	18	53566471	rs2202572	C	A	0.67	0.68	0.58
Lung function	*QTRT1*	19	10819967	rs11085744	T	C	0.56	0.59	0.66
Lung function	*TSHZ3*	19	31829613	rs9636166	C	A	0.14	0.12	0.09
Lung function	*ZFP82*	19	36881643	rs2967516	A	G	0.72	0.71	0.7
Lung function	*LTBP4*	19	41117300	rs34093919	G	A	0.99	0.99	NA
COPD	*CYP2A6*	19	41339896	rs12459249	C	T	0.68	0.71	0.68
COPD	*DMWD*	19	46294136	rs72626215	G	A	0.73	0.75	0.89
Lung function	*BMP2*	20	6626218	rs2145272	G	A	0.36	0.36	0.34
Lung function	*JAG1*	20	10745545	rs6032942	G	C	0.76	0.76	0.84
Lung function	*ABHD12*	20	25282608	rs2236180	C	T	0.19	0.2	0.07
Lung function	*KIF3B*	20	30858967	rs4413223	A	G	0.17	0.17	0.5
Lung function	*GDF5*	20	34025756	rs143384	A	G	0.58	0.58	0.13
Lung function	*EYA2*	20	45486817	rs12481092	T	C	0.26	0.27	0.12
Lung function	*SLC2A4RG*	20	62372706	rs4809221	A	G	0.7	0.72	0.71
Lung function	*MRPS6*	21	35368402	rs12627254	G	T	0.87	0.87	0.96
COPD	*KCNE2*	21	35661745	rs2096468	A	C	0.44	0.44	0.6
Lung function	*KCNE2*	21	35675966	rs62213732	C	T	0.37	0.37	0.52
Lung function	*MICAL3*	22	18448113	rs1978968	T	C	0.22	0.21	0.09
COPD	*MICAL3*	22	18488883	rs9617650	C	G	0.2	0.19	0.2
Lung function	*SCARF2*	22	20790723	rs9610955	C	G	0.18	0.19	0.05
Lung function	*MN1*	22	28181399	rs2283847	T	C	0.55	0.56	0.34

COPD	*SYN3*	22	33335386	rs73158393	G	C	0.26	0.24	0.06
Lung function	*PPP6R2*	22	50867711	rs113111175	T	C	0.11	0.1	0.03

Alt. allele=Alternative allele; NHW=Non-Hispanic Whites; AA=African Americans; ECLIPSE= Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points

Supplementary Table 1.2. SNPs associated with annual change in emphysema at the suggestive significance (P < 1e-05) in Meta-analysis of all subjects

						COPDGene NHW				COPDGene AA				ECLIPSE Whites				Meta-analysis of All subjects		
Nearest gene	Chr.	Position	SNP	Risk allele	Alt.allele	RAF	Beta	SE	P	RAF	Beta	SE	P	RAF	Beta	SE	P	Beta	SE	P
Change in %LAA-950																				
WWC1	5	167724186	rs13164530	T	G	0.136	0.132	0.035	1.96E-04	0.068	0.172	0.079	3.01E-02	0.144	0.144	0.054	0.00761	0.14	0.028	4.32E-07
KLK15	19	51345568	rs2659051	G	C	0.779	0.128	0.03	2.62E-05	0.821	0.163	0.059	5.52E-03	0.786	0.053	0.047	0.25447	0.115	0.023	8.92E-07
SP8	7	21028614	rs73269804	A	C	0.925	0.134	0.047	4.07E-03	0.722	0.148	0.043	6.77E-04	0.929	0.147	0.072	0.04261	0.142	0.029	9.92E-07
PRR25	16	867299	rs79237026	A	C	0.024	0.171	0.088	5.16E-02	0.087	0.263	0.074	3.55E-04	0.02	0.565	0.174	0.00121	0.258	0.054	1.58E-06
NSD2	4	1960600	rs7377981	C	T	0.927	0.126	0.048	8.12E-03	0.981	0.382	0.156	1.44E-02	0.936	0.291	0.077	0.00017	0.185	0.039	2.43E-06
RBMS3	3	29338495	rs6799108	C	T	0.082	0.098	0.044	2.75E-02	0.172	0.138	0.051	6.98E-03	0.08	0.242	0.068	0.00039	0.14	0.03	3.30E-06
ALDH1B1	9	38395940	rs2228094	T	C	0.04	0.183	0.065	4.72E-03	0.122	0.163	0.063	9.93E-03	0.04	0.279	0.103	0.00667	0.19	0.041	4.36E-06
CBFA2T3	16	88961129	rs116233084	G	T	0.016	0.056	0.098	5.69E-01	0.075	0.412	0.081	4.42E-07	0.018	0.236	0.147	0.10823	0.263	0.058	4.97E-06
PLXNA2	1	208796922	rs115506418	A	G	0.91	0.15	0.043	4.90E-04	0.93	0.093	0.079	2.41E-01	0.91	0.187	0.067	0.00518	0.149	0.033	5.76E-06
MDGA2	14	48377425	rs144251028	T	G	0.975	0.214	0.08	7.28E-03	0.92	0.282	0.073	1.15E-04	0.98	0.049	0.132	0.70836	0.223	0.05	8.00E-06
TXNRD1	12	104612225	rs12303096	A	G	0.944	0.112	0.052	3.25E-02	0.764	0.162	0.045	3.59E-04	0.942	0.14	0.08	0.08108	0.141	0.032	8.13E-06
OR2Y1	5	180142163	rs7717059	T	G	0.726	0.084	0.027	2.23E-03	0.424	0.093	0.04	1.95E-02	0.731	0.094	0.042	0.02455	0.088	0.02	8.37E-06
TIAM1	21	32627826	rs2833351	C	G	0.916	0.135	0.044	2.22E-03	0.665	0.129	0.042	2.23E-03	0.91	0.08	0.065	0.21456	0.122	0.028	8.65E-06
PMCH	12	102684915	rs146354672	G	A	0.987	0.262	0.11	1.74E-02	0.989	0.569	0.211	7.26E-03	0.987	0.53	0.172	0.00215	0.377	0.085	9.30E-06
CREB5	7	28555835	rs216707	T	A	0.349	0.09	0.027	7.44E-04	0.432	0.062	0.041	1.33E-01	0.329	0.107	0.042	0.01037	0.088	0.02	9.46E-06
PTPRD	9	8975175	rs10977419	C	T	0.102	0.179	0.041	1.24E-05	0.024	0.102	0.135	4.50E-01	0.098	0.081	0.063	0.19967	0.147	0.033	9.63E-06
C11orf74	11	36738257	rs11033758	A	G	0.568	0.088	0.025	3.65E-04	0.908	0.135	0.067	4.38E-02	0.575	0.075	0.041	0.06759	0.089	0.02	9.74E-06
ZHX2	8	123674204	rs68133336	G	A	0.571	0.063	0.025	1.25E-02	0.747	0.086	0.046	6.18E-02	0.564	0.136	0.039	0.00048	0.085	0.019	9.88E-06
Change in ALD																				
WEE1	11	9593574	rs7940672	A	G	0.157	-0.096	0.033	3.70E-03	0.152	-0.139	0.053	8.88E-03	0.148	-0.148	0.054	0.0059	-0.116	0.025	2.74E-06
TSHZ2	20	51935555	rs200649	T	C	0.825	-0.121	0.032	1.47E-04	0.904	-0.068	0.065	2.92E-01	0.841	-0.132	0.051	0.00957	-0.116	0.025	3.32E-06
MYT1L	2	1806778	rs11893101	A	G	0.027	-0.229	0.076	2.54E-03	0.082	-0.265	0.074	3.29E-04	0.024	-0.104	0.128	0.41434	-0.227	0.049	3.41E-06
NDUFA10	2	240866757	rs61285275	A	G	0.086	-0.111	0.043	9.37E-03	0.333	-0.172	0.04	1.86E-05	0.102	-0.021	0.062	0.73874	-0.122	0.026	4.43E-06

FAM135B	8	138569304	rs148752344	T	C	0.08	-0.163	0.045	2.72E-04	0.094	-0.09	0.067	1.79E-01	0.074	-0.193	0.072	0.00775	-0.151	0.033	4.71E-06
GNAI1	7	79818957	rs28435776	A	G	0.823	-0.118	0.032	1.96E-04	0.947	-0.142	0.087	1.02E-01	0.829	-0.107	0.051	0.03474	-0.117	0.026	4.75E-06
FRK	6	115618884	rs12190538	A	G	0.697	-0.088	0.026	8.37E-04	0.741	-0.125	0.042	2.86E-03	0.691	-0.057	0.04	0.15286	-0.089	0.02	5.18E-06
RIMBP2	12	131226142	rs10848195	G	A	0.446	-0.091	0.024	2.11E-04	0.334	-0.093	0.039	1.77E-02	0.456	-0.053	0.038	0.16675	-0.083	0.018	5.85E-06
SCEL	13	78032021	rs17067745	A	G	0.967	-0.252	0.072	4.77E-04	0.858	-0.183	0.058	1.53E-03	0.972	-0.046	0.125	0.71109	-0.191	0.042	6.34E-06
BRINP1	9	122304582	rs17578200	G	C	0.135	-0.115	0.037	1.70E-03	0.041	-0.068	0.116	5.57E-01	0.141	-0.198	0.058	0.00059	-0.135	0.03	6.82E-06
RTL1	14	101310879	rs12587062	G	A	0.763	-0.119	0.029	3.52E-05	0.911	-0.045	0.067	5.05E-01	0.784	-0.087	0.046	0.05772	-0.102	0.023	7.56E-06
IGSF9B	11	133843315	NA	G	A	0.308	-0.098	0.03	1.12E-03	0.367	-0.074	0.042	7.51E-02	0.324	-0.12	0.047	0.01083	-0.096	0.022	8.67E-06
LRRC4C	11	40677144	rs11035996	T	C	0.802	-0.072	0.03	1.72E-02	0.831	-0.195	0.049	6.69E-05	0.803	-0.08	0.046	0.08459	-0.1	0.022	8.70E-06
NXN	17	874466	rs113611215	G	A	0.962	-0.183	0.063	3.87E-03	0.947	-0.334	0.09	2.24E-04	0.964	-0.099	0.101	0.32919	-0.205	0.046	8.84E-06
MCC	5	112824166	rs348941	C	G	0.693	-0.111	0.026	2.36E-05	0.638	-0.062	0.039	1.11E-01	0.69	-0.048	0.04	0.23764	-0.085	0.019	8.98E-06
ZNF770	15	35312826	rs4924470	C	T	0.089	-0.092	0.042	3.13E-02	0.284	-0.133	0.041	1.38E-03	0.084	-0.161	0.068	0.01715	-0.12	0.027	9.04E-06
RNF150	4	141828469	rs2303414	G	A	0.989	-0.301	0.117	9.95E-03	0.989	-0.393	0.206	5.67E-02	0.988	-0.586	0.175	0.00085	-0.39	0.088	9.31E-06

%LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density; RAF=Risk allele frequency

Supplementary Table 1.3. SNPs associated with annual change in emphysema at the suggestive significance (P < 1e-05) in Meta-analysis of Whites

						COPDGene NHW				ECLIPSE Whites				Meta-analysis of Whites			COPDGene AA			
Nearest gene	Chr.	Position	SNP	Risk allele	Alt. allele	RAF	Beta	SE	P	RAF	Beta	SE	P	Beta	SE	P	RAF	Beta	SE	P
Change in %LAA-950																				
NXPH2	2	139581765	rs115047317	G	A	0.969	0.375	0.073	2.55E-07	0.96	0.198	0.101	4.97E-02	0.314	0.059	9.31E-08	NA	NA	NA	NA
VSTM2A	7	54177112	rs1965517	A	G	0.929	0.186	0.048	1.11E-04	0.935	0.27	0.076	4.11E-04	0.21	0.041	2.44E-07	0.786	0.005	0.047	9.23E-01
TENM3	4	182163057	rs78796196	C	A	0.982	0.293	0.094	1.80E-03	0.977	0.584	0.131	9.09E-06	0.391	0.076	2.86E-07	NA	NA	NA	NA
PDIK1L	1	26439829	rs183758663	A	C	0.964	0.251	0.072	4.69E-04	0.966	0.417	0.112	2.17E-04	0.299	0.06	7.55E-07	NA	NA	NA	NA
CATSPER4	1	26519293	rs56177125	G	A	0.959	0.269	0.062	1.33E-05	0.962	0.227	0.1	2.29E-02	0.258	0.052	9.21E-07	NA	NA	NA	NA
CNKSR1	1	26507612	rs35600413	A	T	0.967	0.294	0.068	1.65E-05	0.968	0.234	0.108	3.03E-02	0.277	0.058	1.54E-06	NA	NA	NA	NA
RAB3A	19	18314266	rs73001430	C	A	0.959	0.277	0.068	5.08E-05	0.956	0.233	0.104	2.53E-02	0.264	0.057	3.82E-06	NA	NA	NA	NA
SGCD	5	154996354	rs4958830	A	T	0.025	0.267	0.083	1.28E-03	0.033	0.399	0.116	6.11E-04	0.311	0.067	3.85E-06	NA	NA	NA	NA
WWC1	5	167724186	rs13164530	T	G	0.136	0.132	0.035	1.96E-04	0.144	0.144	0.054	7.61E-03	0.136	0.03	4.56E-06	0.068	0.172	0.079	3.01E-02
MSI2	17	55671179	rs76716491	T	G	0.012	0.461	0.121	1.37E-04	0.011	0.463	0.186	1.31E-02	0.462	0.101	5.21E-06	NA	NA	NA	NA
PLXNA2	1	208806940	rs6685613	T	C	0.863	0.142	0.036	6.37E-05	0.869	0.126	0.059	3.17E-02	0.138	0.03	5.64E-06	0.749	0.011	0.045	8.15E-01
LOXL4	10	100053881	rs75822602	T	C	0.033	0.321	0.071	5.81E-06	0.038	0.143	0.102	1.61E-01	0.263	0.058	5.88E-06	NA	NA	NA	NA
NECTIN2	19	45383037	rs11673139	A	T	0.909	0.166	0.043	1.28E-04	0.906	0.158	0.067	1.80E-02	0.164	0.036	6.59E-06	0.909	0.035	0.077	6.50E-01
DCAF6	1	168028674	rs11803859	G	A	0.927	0.183	0.048	1.38E-04	0.925	0.169	0.071	1.70E-02	0.178	0.04	6.76E-06	0.973	0.033	0.127	7.95E-01
SEC16B	1	177721831	rs12131792	A	G	0.038	0.285	0.064	9.63E-06	0.04	0.139	0.097	1.53E-01	0.241	0.054	7.28E-06	NA	NA	NA	NA
NDUFA4	7	10572624	rs149165422	A	G	0.984	0.398	0.101	8.91E-05	0.985	0.338	0.158	3.27E-02	0.38	0.085	8.32E-06	NA	NA	NA	NA
CENPC	4	67813755	rs1481266	C	T	0.028	0.217	0.082	7.91E-03	0.026	0.522	0.127	4.52E-05	0.306	0.069	8.74E-06	NA	NA	NA	NA
ZNF839	14	102785629	rs150789694	A	C	0.985	0.49	0.114	1.85E-05	0.987	0.281	0.189	1.37E-01	0.434	0.098	9.01E-06	NA	NA	NA	NA
Change in ALD																				
LRIG2	1	113591452	rs146580149	A	G	0.951	-0.198	0.061	1.21E-03	0.944	-0.34	0.087	1.05E-04	-0.244	0.05	1.04E-06	NA	NA	NA	NA
DR1	1	93814700	rs116807672	C	T	0.982	-0.497	0.105	2.24E-06	0.981	-0.26	0.157	9.76E-02	-0.424	0.087	1.17E-06	NA	NA	NA	NA
FNBP1L	1	93981950	rs10518536	A	G	0.981	-0.482	0.103	3.09E-06	0.979	-0.264	0.152	8.36E-02	-0.414	0.085	1.30E-06	NA	NA	NA	NA
LY96	8	74983233	rs74859713	T	C	0.03	-0.344	0.074	3.05E-06	0.033	-0.162	0.113	1.50E-01	-0.29	0.062	2.56E-06	0.026	0.088	0.146	5.45E-01
FAM135B	8	138579554	rs72728184	T	C	0.089	-0.173	0.042	4.58E-05	0.084	-0.156	0.068	2.13E-02	-0.168	0.036	2.83E-06	0.06	0.014	0.082	8.60E-01

FRK	6	116002717	rs12205186	A	T	0.686	-0.113	0.026	1.08E-05	0.686	-0.07	0.04	8.53E-02	-0.1	0.022	3.42E-06	0.674	-0.028	0.04	4.89E-01
GZMB	14	25120382	rs7154975	G	A	0.531	-0.081	0.024	6.87E-04	0.537	-0.126	0.038	9.51E-04	-0.094	0.02	3.49E-06	0.726	-0.002	0.043	9.72E-01
RNF144B	6	18575300	rs6916983	T	C	0.725	-0.09	0.026	6.40E-04	0.721	-0.132	0.041	1.42E-03	-0.102	0.022	4.18E-06	0.608	0.022	0.038	5.61E-01
TSHZ2	20	51935555	rs200649	T	C	0.825	-0.121	0.032	1.47E-04	0.841	-0.132	0.051	9.57E-03	-0.124	0.027	4.26E-06	0.904	-0.068	0.065	2.92E-01
CCDC18	1	93675574	rs115453772	C	G	0.985	-0.463	0.112	3.93E-05	0.984	-0.347	0.171	4.22E-02	-0.428	0.094	5.18E-06	NA	NA	NA	NA
MSI2	17	55671179	rs76716491	T	G	0.012	-0.388	0.119	1.09E-03	0.011	-0.613	0.185	9.60E-04	-0.454	0.1	5.67E-06	NA	NA	NA	NA
ADAMTSL1	9	18806785	rs72690583	T	C	0.01	-0.327	0.12	6.36E-03	0.012	-0.702	0.174	5.86E-05	-0.447	0.099	5.80E-06	NA	NA	NA	NA
LIMCH1	4	41690410	rs16853485	C	T	0.826	-0.109	0.032	5.38E-04	0.83	-0.146	0.049	2.99E-03	-0.12	0.026	6.11E-06	0.812	0.008	0.048	8.66E-01
RTL1	14	101310879	rs12587062	G	A	0.763	-0.119	0.029	3.52E-05	0.784	-0.087	0.046	5.77E-02	-0.11	0.024	6.14E-06	0.911	-0.045	0.067	5.05E-01
PTPRK	6	128831588	rs150582742	C	T	0.034	-0.361	0.075	1.68E-06	0.038	-0.109	0.111	3.28E-01	-0.282	0.062	6.17E-06	NA	NA	NA	NA
ZNF600	19	53280252	rs10414169	A	T	0.982	-0.322	0.103	1.69E-03	0.984	-0.645	0.18	3.37E-04	-0.402	0.089	6.43E-06	0.786	-0.018	0.054	7.33E-01
KSR2	12	118176064	rs4767602	C	T	0.106	-0.179	0.041	1.32E-05	0.112	-0.097	0.061	1.13E-01	-0.154	0.034	6.50E-06	0.055	0.009	0.092	9.25E-01
BRINP1	9	122304582	rs17578200	G	C	0.135	-0.115	0.037	1.70E-03	0.141	-0.198	0.058	5.94E-04	-0.139	0.031	6.79E-06	0.041	-0.068	0.116	5.57E-01
SNCA	4	90718390	rs33978842	G	C	0.96	-0.232	0.064	2.69E-04	0.964	-0.281	0.106	8.23E-03	-0.245	0.055	7.16E-06	NA	NA	NA	NA
IFT74	9	27002194	rs145997721	C	A	0.035	-0.267	0.068	9.86E-05	0.032	-0.251	0.113	2.68E-02	-0.263	0.059	7.34E-06	NA	NA	NA	NA

%LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density; Chr.=Chromosome; Alt. Allele=Alternative Allele; RAF=Risk allele frequency; NHW=Non-Hispanic Whites; AA=African Americans; ECLIPSE= Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points

Supplementary Table 1.4. Association of *DSP* variant, rs2076295, with annual change in emphysema stratified by COPD status and presence of emphysema in European ancestry (Transformed measure)

	COPDGene NHW				ECLIPSE Whites				Meta-analysis of Whites		
Sub-group	**N**	**Beta**	**SE**	**P**	**N**	**Beta**	**SE**	**P**	**Beta**	**SE**	**P**
Change in %LAA-950											
Overall	3030	0.0842	0.0249	7.50E-04	1397	0.092	0.0386	1.72E-02	0.0865	0.021	3.66E-05
COPD case	1030	0.0558	0.0439	2.04E-01	1251	0.1027	0.0407	1.16E-02	0.081	0.0298	6.58E-03
COPD control	1363	0.0888	0.0368	1.60E-02	146	-0.0642	0.1202	5.94E-01	0.0757	0.0352	3.15E-02
Emphysema case	1041	0.1072	0.0423	1.14E-02	1112	0.0821	0.043	5.67E-02	0.0949	0.0302	1.66E-03
Emphysema control	1989	0.0786	0.0306	1.04E-02	285	0.2256	0.0894	1.22E-02	0.094	0.029	1.17E-03
Change in ALD											
Overall	3030	-0.0611	0.0245	1.27E-02	1397	-0.0629	0.0385	1.02E-01	-0.0617	0.0207	2.85E-03
COPD case	1030	-0.0449	0.043	2.96E-01	1251	-0.0613	0.0405	1.30E-01	-0.0536	0.0295	6.89E-02
COPD control	1363	-0.0603	0.0364	9.79E-02	146	-0.0867	0.1209	4.74E-01	-0.0625	0.0348	7.30E-02
Emphysema case	1041	-0.1184	0.0412	4.14E-03	1112	-0.0355	0.0428	4.07E-01	-0.0785	0.0297	8.16E-03
Emphysema control	1989	-0.0455	0.0306	1.37E-01	285	-0.1855	0.0915	4.37E-02	-0.0596	0.029	4.01E-02

Outcome inversely normal transformed; %LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density; NHW=Non-Hispanic Whites; ECLIPSE= Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points

Supplementary Table 1.5. Association of *DSP* variant, rs2076295, with annual change in emphysema stratified by COPD status and presence of emphysema in European ancestry (Untransformed measure)

	COPDGene NHW				ECLIPSE Whites				Meta-analysis of Whites		
Sub-group	**N**	**Beta**	**SE**	**P**	**N**	**Beta**	**SE**	**P**	**Beta**	**SE**	**P**
Change in %LAA-950											
Overall	3030	0.0523	0.0179	3.57E-03	1397	0.1975	0.0821	1.63E-02	0.0589	0.0175	7.74E-04
COPD case	1030	0.0503	0.0424	2.35E-01	1251	0.231	0.0904	1.07E-02	0.0829	0.0384	3.08E-02
COPD control	1363	0.0381	0.0172	2.69E-02	146	-0.1247	0.1229	3.12E-01	0.035	0.017	4.00E-02
Emphysema case	1041	0.115	0.0472	1.50E-02	1112	0.1971	0.0996	4.82E-02	0.13	0.0427	2.31E-03
Emphysema control	1989	0.0232	0.0107	3.05E-02	285	0.2279	0.1004	2.40E-02	0.0255	0.0107	1.67E-02
Change in ALD											
Overall	3030	-0.1311	0.0502	9.14E-03	1397	-0.26	0.1402	6.40E-02	-0.1457	0.0473	2.06E-03
COPD case	1030	-0.1078	0.0856	2.08E-01	1251	-0.2503	0.1503	9.60E-02	-0.1427	0.0744	5.51E-02
COPD control	1363	-0.1131	0.0736	1.25E-01	146	-0.3673	0.3699	3.23E-01	-0.1228	0.0722	8.91E-02
Emphysema case	1041	-0.2716	0.0866	1.75E-03	1112	-0.1696	0.1447	2.41E-01	-0.2447	0.0743	9.87E-04
Emphysema control	1989	-0.0869	0.0611	1.55E-01	285	-0.7012	0.4059	8.52E-02	-0.1005	0.0604	9.61E-02

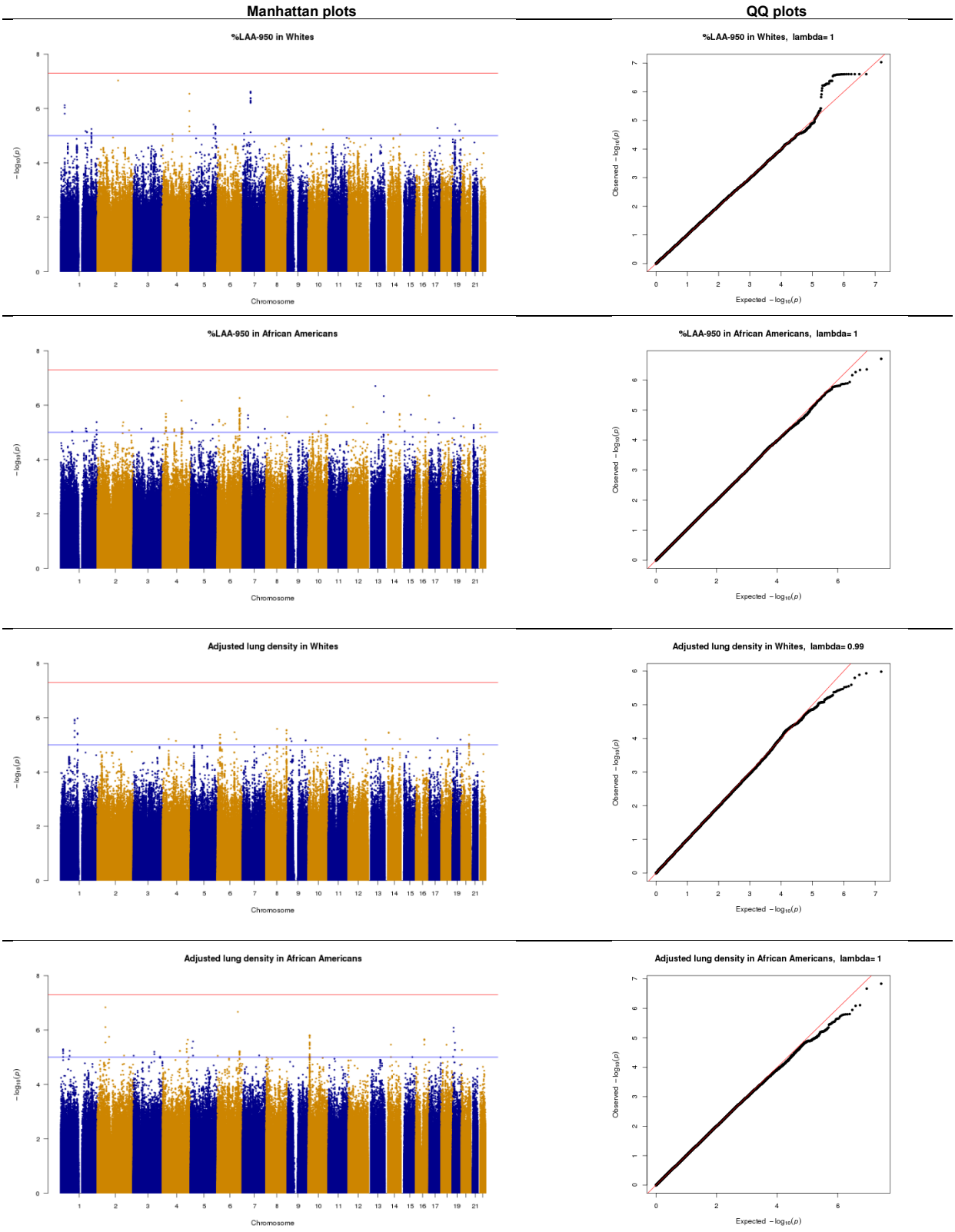
Outcome untransformed; %LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density; NHW=Non-Hispanic Whites; ECLIPSE= Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points

Supplementary Table 1.6. Association of lung function polygenic risk score with baseline emphysema and annual change in emphysema (Untransformed measure)

	Baseline emphysema			Annual change in emphysema		
	Beta	**SE**	**P**	**Beta**	**SE**	**P**
%LAA-950						
Meta-Analysis of Europeans	0.082	0.0109	6.16E-14	0.0012	0.001	2.13E-01
COPDGene NHW	0.0793	0.0119	3.53E-11	0.0008	0.001	4.45E-01
ECLIPSE White	0.0962	0.0273	4.30E-04	0.0112	0.0047	1.82E-02
COPDGene AA	0.0595	0.0174	6.47E-04	0.0001	0.0016	9.69E-01
ALD						
Meta-Analysis of Europeans	-0.16	0.0241	3.14E-11	-0.0055	0.0027	4.16E-02
COPDGene NHW	-0.159	0.028	1.45E-08	-0.0054	0.0029	6.16E-02
ECLIPSE White	-0.1629	0.0474	6.09E-04	-0.0067	0.0081	4.10E-01
COPDGene AA	-0.0956	0.0556	8.58E-02	0.0046	0.006	4.44E-01

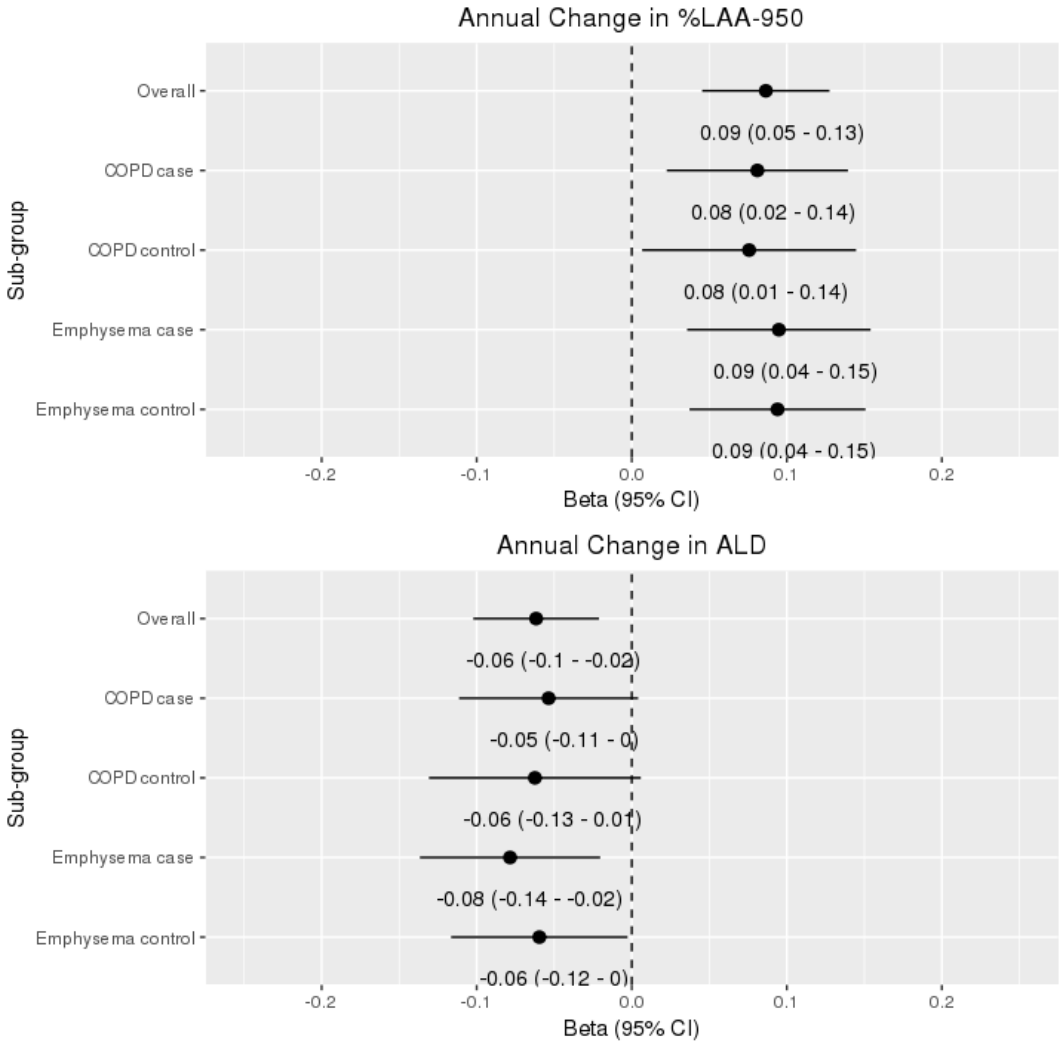
Outcome untransformed; %LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density; NHW=Non-Hispanic Whites; ECLIPSE= Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points

Supplementary Figure 1.1. Manhattan plots and quantile-quantile (Q-Q) plots of association results for annual change in emphysema



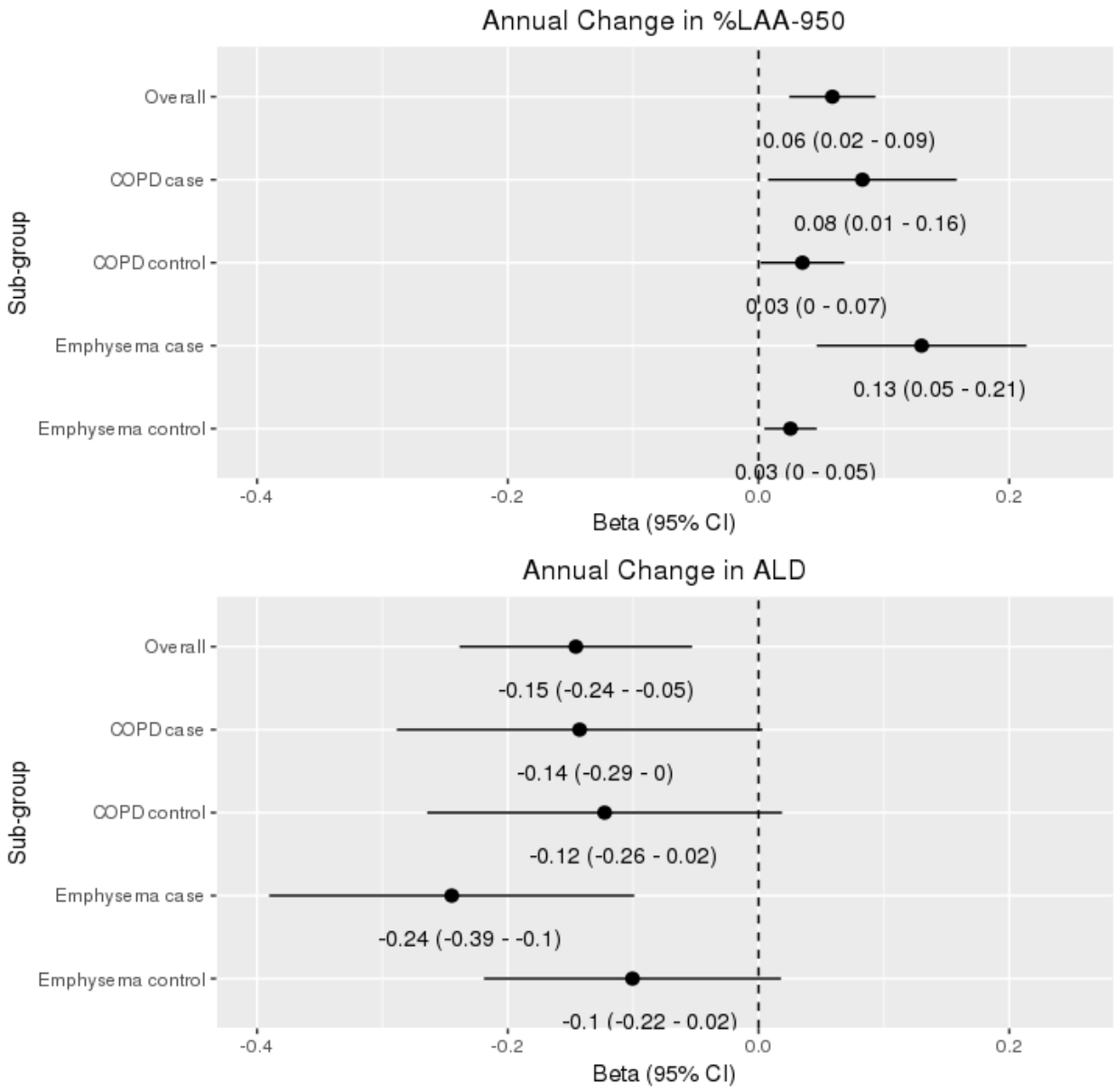
%LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density

Supplementary Figure 1.2. Forest plot of *DSP* variant association in European ancestry (Transformed measure)



%LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density

Supplementary Figure 1.3. Forest plot of *DSP* variant association in European ancestry (Untransformed measure)



%LAA-950=percentage of low-attenuation area less than -950 Hounsfield units; ALD=Adjusted lung density

Chapter 3. Genome-wide gene-by-smoking interaction study of Chronic Obstructive Pulmonary Disease

Chapter 3. Genome-wide gene-by-smoking interaction study of Chronic Obstructive Pulmonary Disease

Introduction

Risk for Chronic Obstructive Pulmonary Disease (COPD) is determined by both cigarette smoking and genetic susceptibility. The adverse effects of smoking on risk of COPD may differ by an individual's genetic susceptibility, which raises the potential gene-by-smoking interactions. However, little is known about gene-by-smoking interactions on COPD risk.

It has been reported that SERPINA1 variants, PI*Z allele and rs800738, interact with cigarette smoking on spirometric measure of lung function [1,2] and COPD risk [3], respectively. For COPD-related traits, genome-wide gene-by-smoking interaction studies have focused on quantitative measures of lung function [4,5]. While spirometric measures of lung function are used to diagnose COPD, no genome-wide studies have investigated gene-by-smoking interaction on risk of COPD itself.

A recent large-scale genome-wide association study (GWAS) identified 82 distinct loci associated with risk of COPD [6]. However, these identified variants explained less than 10% of the phenotypic variability on the liability scale. To fill the gap of the genetics of COPD explained by common variants, more of the phenotypic variability might be explained by including gene-by-smoking interactions in GWAS model.

A major challenge of studying gene-by-environment interactions is the much larger sample size required compared to conventional GWAS to detect marginal effects of genes [7]. The 2-degree-

of-freedom (2df) joint test leverages genetic main effects and gene-by-environment interaction effects simultaneously and can provide better power than the standard interaction test, 1-degree-of-freedom (1df) test [8]. Using a 2df test, recent large-scale genome-wide gene-by-environment interaction studies of complex traits have identified new genetic factors as well as gene-by-environment interactions [4,9–13]. The availability of the large-scale UK Biobank (UKB) study, collecting a wide range of phenotypes as well as genetic data, could provide a promising opportunity to detect gene-by-environment interactions.

Here, we performed genome-wide gene-by-smoking interaction analyses of COPD in the UKB study to identify novel genetic variants for risk of COPD while accounting for potential smoking interactions and assessed the impact of gene-by-smoking interactions on risk of COPD at known COPD and lung function GWAS loci.

Methods

Study populations

The UKB is a population-based cohort of volunteers where over 500,000 individuals were originally recruited [14]. We used UKB subjects as our discovery set. We also used two additional datasets, the COPDGene Study and the SpiroMeta Consortium, to further investigate significant results from UKB. COPDGene recruited former and current smokers whose smoking history is at least 10 pack-years [15]. SpiroMeta is comprised of a total of 79,055 individuals from 22 studies [16]. All participants provided written informed consent and studies were approved by local Research Ethics Committees and/or Institutional Review boards.

Spirometric measures and genetic data

Details of quality controls (QC) of spirometric measures, genetic markers and subjects in the UKB study are previously described [6,14,16]. Briefly, to determine lung function, measures of forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) were derived from the spirometry volume-time series data, subjected to additional quality control based on ATS/ERS criteria [16,17]. Genotyping was performed using Axiom UK BiLEVE array and Axiom Biobank array (Affymetrix, Santa Clara, California, USA) and imputed to the Haplotype Reference Consortium version 1.1 panel. We included independent subjects of European ancestry based on a combination of self-reported ethnicity data and principal components (PCs) data provided by UKB.

Measures of smoking exposure

We assigned smoking status to individuals in UKB based on their responses on questionnaires. Never-smokers included non-current-smokers or those who smoked less than 100 cigarettes in their life. Ever-smokers were defined as either current, most days (current or all days in the past) or smoked occasionally.

For genome-wide gene-by-smoking interaction testing, we considered 2 binary smoking variables: Ever/Never and Current/Non-current smoker. For Ever/Never smokers, former- and current-smokers were included into the ever-smoker group. For Current/Non-current smokers, former- and never- smokers were included into the non-current-smoker group. Smoking variables were coded as 0 and 1 for the unexposed and exposed groups, respectively. Here, we refer Ever/Never-smoker analysis as “GxEver-smoking analysis” and Current/Non-current smoker analysis as “GxCurrent-smoking analysis”.

Outcome

We defined COPD cases based on pre-bronchodilator spirometry according to modified Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria for moderate airflow limitation: FEV_1 less than 80% of predicted value (using reference equations from [18]), and the ratio of FEV_1/FVC less than 0.7.

Genetic analysis

We included markers with minor allele frequency (MAF) ≥ 0.01 and imputation quality score (r^2) ≥ 0.5. We performed a logistic regression analysis considering both genetic main effect and gene-by-smoking interaction effect (2df joint test) in the genome-wide analysis, adjusting for age, sex, genotyping array and the first 10 PCs. We used a 2df test to jointly assess both genetic main effect and interaction effect to search for new genetic variant of COPD and a 1df SNP-by-smoking interaction test to assess the interaction effects alone. Additionally, marginal GWAS were run stratified by each smoking variable. All genome-wide analyses were performed using Plink software (version 2.0, www.cog-genomics.org/plink/2.0).

Conditional analysis

We defined distinct 'loci' using a 1-Mb window (+/- 500kb) around the most significant SNP (lead variant). As our joint analysis was likely to include substantial overlap with previously described association studies of main effects for COPD, we performed conditional analysis of lead variants to determine whether our signals were independent of known loci for COPD [6] or lung function [16]. As the current GCTA (http://cnsgenomics.com/software/gcta) tool does not account for gene-by-environment interactions in their conditional analysis, we took a stratified approach. We stratified by smoking exposed- and unexposed- groups and conditioned on recognized SNPs from previous GWASs of COPD [6] or lung function [16] within 2-Mb of the lead

variants. The conditioned 2df test for genetic main effects and interaction effects was then calculated on the conditioned stratified results using the following equations [10,19]. For the 1df test,

$$\mathrm{Z} = \frac{\gamma_G^{(1)} - \gamma_G^{(0)}}{\sqrt{SE(\gamma_G^{(1)})^2 + SE(\gamma_G^{(0)})^2 - 2rSE(\gamma_G^{(1)})SE(\gamma_G^{(0)})}}$$

where $\gamma_G^{(1)}$ and $\gamma_G^{(0)}$ represent stratum-specific genetic effects; $SE(\gamma_G^{(1)})$ and $SE(\gamma_G^{(0)})$ are their respective standard error (SE); and r is the Spearman rank correlation coefficient between $\gamma_G^{(1)}$ and $\gamma_G^{(0)}$, calculated from the genome-wide results. The Z statistics approximately follows a standard normal distribution under $H_0: \beta_{GE} = 0$. For the 2df test,

$$\mathrm{X} = \left[\frac{\gamma_G^{(1)}}{SE(\gamma_G^{(1)})}\right]^2 + \left[\frac{\gamma_G^{(0)}}{SE(\gamma_G^{(0)})}\right]^2$$

which approximately follows a 2df chi-squared distribution under $H_0: \beta_G = \beta_{GE} = 0$ when the two strata are independent.

Dose response analysis

To further characterize significant results, we conducted a dose response analysis in all subjects and ever smokers (a secondary analysis). We tested gene-by-smoking dose interaction using the standard 1df test. We considered three quantitative measures of smoking dose: smoking duration, pack-years (PY) and cigarettes per day (CPD). We considered both a quantitative variable and a categorical variable grouped based on quartiles.

Replication

As the COPDGene cohort is enriched for heavy smokers, we hypothesized SNPs presenting a stronger association among the exposed group in UKB would also show some marginal

associations with risk of COPD in COPDGene subjects. With the selected SNPs, we tested for a marginal association between SNP and risk of COPD, adjusting for age, sex, smoking status, pack-years and genetic ancestry PCs in 5,342 Non-Hispanic White (NHW) subjects from COPDGene. With the selected SNPs, we further tested gene-by-smoking dose interaction based on the 1df test. As the Fagerstrom Test for Nicotine Dependence (FTND) measure was collected for current smokers in COPDGene, we additionally tested gene-by-FTND interaction. We considered both a quantitative variable of FTND score and a categorical variable grouped into mild (0-3), moderate (4-6) and severe (7-10) [20].

We also attempted to replicate our results by lookup in a GWAS analysis of spirometric measures of lung function (FEV_1, FVC and FEV_1/FVC) stratified by ever- and never-smoker groups in SpiroMeta. Using summary statistics of these stratified results, we derived statistics of a 1df interaction test and a 2df joint test based on the same approach used for the stratified conditional analysis [10,19]. The analysis in SpiroMeta has been previously published [16]. Briefly, each study was performed as a linear regression adjusting for age, age^2, sex, and height, by using rank-based inverse normal transformation, adjusting for population substructure by including genetic ancestry PCs or as linear mixed models, and performing separate analyses for ever- and never-smokers or using a covariate for smoking (for studies of related subjects). Results were combined by using a fixed-effects meta-analysis.

Results

Subject characteristics

We analyzed 200,766 subjects including 179,689 controls and 21,077 COPD cases in the UKB study (**Table 2.1**). UKB subjects included 71,591 ever- (current- and former-smokers combined)

and 129,175 never-smokers; and 14,590 current- and 186,176 non-current-smokers (never- and former-smokers combined). COPDGene subjects included NHW 3,361 former-smokers and 1,981 current-smokers. While UKB subjects had a higher proportion of COPD cases among current-smokers (31.5%) than former- (13.8%) and never-smokers (6.7%), COPDGene subjects (who were enriched for moderate-to-severe COPD) showed a higher proportion of COPD cases among former-smokers (54.5%) than current-smokers (49.4%).

Genome-wide Results

The analysis workflow is depicted in **Figure 2.1**.

2df joint test. We identified 48 loci for GxEver- and 55 loci for GxCurrent- smoking analysis (defined using 1-Mb windows) at genome-wide significance (P<5.00E-08) (**Supplemental Figure 2.1**). The lead variants of 15 loci for GxEver- and 19 loci for GxCurrent- smoking analysis were previously reported variants in GWAS of COPD or lung function [6,16]. For the remaining loci, we conducted a conditional analysis to search for new signal (see **Method**). After adjusting for reported variants, 2 loci, 16q22.1 - *SMPD3* (lead variant: rs141322661, P_{2df} =3.92E-09 from GxEver- and P_{2df} =1.45E-08 from GxCurrent- smoking analysis) and 19q13.2 - *EGLN2* (lead variant: rs2604894, P_{2df} =5.87E-09 from GxCurrent smoking analysis), maintained genome-wide significance (**Table 2.2 and Figure 2.2**).

In previous UKB GWAS of COPD examining only marginal effects [6], rs141322661 at 16q22.1 reached genome-wide significance (P=1.88E-09), but not in overall meta-analysis of UKB and International COPD Genetics Consortium (ICGC) (P=1.90E-08) and rs2604894 at 19q13.2 did not reach genome-wide significance (P=1.17E-04). In our current/non-current stratified GWAS, rs2604894 showed a stronger association among current-smokers (OR (95% CI)=0.86 (0.82-

0.91), P=2.52E-07) compared to among non-current-smokers (OR (95% CI)=0.96 (0.94-0.99), P=2.77E-03). Other signals were attenuated and did not reach our significance level, indicating that those findings are not novel.

1df interaction test. We identified one locus - 15q25.1 (defined using 1-Mb windows) at genome-wide significance (P <5.00E-08) for both GxEver- and GxCurrent-smoking analyses (**Table 2.2, Figure 2.3 and Supplemental Figure 2.2**). In the GxEver-smoking analysis, the lead variant, rs12440014 in *CHRNB4*, showed $P_{1df\ interaction}$=8.96E-12, presenting significant association in ever-smokers (OR (95% CI)=0.85 (0.82-0.88), P=3.39E-19), but not in never-smokers.

Interaction of reported variants

We examined possible gene-by-smoking interactions on risk of COPD at 82 known COPD-associated loci, 279 known lung function-associated loci and 2 loci previously reporting smoking interactions on lung function or COPD. Because results from the 2df test for these known loci predominantly showed genetic main effects, we evaluated results from the 1df test for gene-by-smoking interactions at Bonferroni corrected significance.

At known loci of COPD, rs55676755 in *CHRNA3* and rs28534575 in *CHRNB4*, significantly interacted with smoking, presenting significant associations in ever-smokers (rs55676755: OR (95% CI)=1.19 (1.15-1.22), P=8.74E-28 and rs28534575: OR (95% CI)=0.85 (0.82-0.88), P=2.53E-18), but not in never-smokers. rs7642001 at 3q26.2 – *MECOM* showed a significant current smoking ($P_{1df\ interaction}$=3.65E-04) but not ever smoking ($P_{1df\ interaction}$=3.11E-01) interaction. At known loci for lung function, there was no statistically significant interactions with smoking. At loci previously reporting smoking interactions, PI*Z allele (rs28929474) - *SERPINA1* significantly

interacted with both ever smoking ($P_{1df\ interaction}$=6.70E-04) and current smoking ($P_{1df\ interaction}$=7.19E-03).

Selected SNPs

To further investigate significant results, we selected 5 SNPs at 5 loci (**Table 2.3** and **Figure 2.4**). In the 2df joint test, rs141322661 at 16q22.1 – *SMPD3* and rs2604894 at 19q13.2 - *EGLN2* reached genome-wide significance, independent of previously described loci of COPD and lung function. In the 1df interaction test, we included rs12440014 at 15q25.1 – *CHRNB4*. In previously reported variants, we selected rs7642001 at 3q26.2 – *MECOM* and rs28929474 - *SERPINA1*, with evidence of interaction.

To examine whether selected SNPs are associated with smoking behavior, we checked regions of selected SNPs in the most recent and largest GWAS of smoking itself [21] (**Supplemental Table 2.1**). 15q25.1 and 19q13.2, were reported to be associated with CPD and current smoking.

To further characterize the interactions, we conducted a dose response analysis (**Supplemental Table 2.2**). We considered smoking duration as a primary measure of smoking dose based on previous reports that smoking duration alone provides larger estimated effects on risk of COPD than PY [22]. rs7642001 at 3q26.2 – *MECOM,* rs28929474 of *SERPINA1*, rs12440014 at 15q25.1 – *CHRNB4* and rs2604894 at 19q13.2 – *EGLN2* exhibited statistically significant interactions with smoking duration on COPD risk (P<0.05). In a dose response analysis within ever-smokers, significances of dose response were attenuated but rs7642001 and rs12440014 were still nominally significant (P<0.05).

Replication

To replicate our findings, we used COPDGene and SpiroMeta (**Table 2.3**).

COPDGene. As COPDGene is enriched for heavy smokers, we hypothesized SNPs showing a stronger association among the exposed group in the UKB should have some marginal associations with COPD risk in COPDGene. rs7642001 at 3q26.2 -*MECOM*, rs28929474 - *SERPINA1,* rs12440014 at 15q25.1 - *CHRNB4* and rs2604894 at 19q13.2 - *EGLN2* were nominally significantly associated with risk of COPD (P<0.05). In a dose response analysis, a stronger association between rs7642001 at 3q26.2 – *MECOM* and COPD was observed with longer duration of smoking ($P_{interaction}$=6.20E-04) (**Supplemental Table 2.3)**. For 1,937 current smokers in COPDGene available for FTND score, rs12440014 at 15q25.1 - *CHRNB4* interacted with higher nicotine dependence ($P_{interaction}$=4.37E-02).

SpiroMeta. We replicated a significant interaction of rs12440014 at 15q25.1 – *CHRNB4* with ever smoking on FEV_1 ($P_{interaction}$<0.05), presenting stronger association among ever-smokers compared to never-smokers in SpiroMeta (**Table 2.3**). We observed a significant interaction for rs7642001 at 3q26.2 - *MECOM* on FEV_1 (P=1.97E-03). However, the direction of interaction effects was opposite between UKB and SpiroMeta. Allele “A” of rs7642001 was more significantly associated with decreased FEV_1 among never-smokers (Beta (95% CI)=-0.04 (-0.05, -0.02), P=3.31E-05) compared to ever-smokers (Beta (95% CI)=-0.01 (-0.03, 0.005), P=1.93E-01), while in UKB more significantly associated with increased risk of COPD among current-smokers (OR (95% CI)=1.20 (1.13-1.27), P=4.54E-10) compared to non-current-smokers (OR (95% CI)=1.07 (1.04-1.09), P=3.09E-07). The stratified analysis for other measures, FEV and FEV_1/FVC, are listed in **Supplemental Table 2.3**.

Discussion

We conducted a genome-wide gene-by-smoking interaction study of COPD risk, using a 2df joint test for genetic main effects and gene-by-smoking interaction effects on COPD risk. Most of the significant signals from 2df test had been reported in previous GWASs of COPD. However, we identified two loci, 16q22.1 – *SMPD3* and 19q13.2 – *EGLN2*, reaching genome-wide significance. We detected a genome-wide significant interaction with smoking at the 15q25.1 locus – *CHRNB4*, previously identified in studies of COPD and smoking behavior. We confirmed *SERPINA1* PI*Z-by-smoking interaction on COPD.

The 16q22.1 region (lead variant: rs141322661) had reached genome-wide significance in previous GWAS of COPD in UKB subjects but not in overall meta-analysis. rs141322661, an intron variant of *SMPD3*, is a low frequent variant in European populations (Allele frequency (G) =0.01 - 0.02), which may make it difficult to replicate this signal. Further investigation of 16q22.1 region is required.

The 19q13.2 region (lead variant: rs2604894) could only be observed through genome-wide association analysis accounting for current-smoking interaction in UKB subjects. In the previous UKB GWAS of COPD not incorporating interaction in the model, rs2604894 did not reach genome-wide significance [6]. In conventional GWAS of COPD using cohorts enriched for smokers, rs2604894 was reported to be significantly associated with COPD risk (OR (95% CI)=0.74 (0.65-0.84), P=3.41E-08) [23]. This study included cohorts such as COPDGene [15] and ECLIPSE [24] designed to identify genetic factors for COPD, recruiting former- and current-smokers. Such different study design between the UKB study, a population-based cohort, and COPD-related cohorts may have attenuated the statistical significance of rs2604894 association and hindered the replication in previous marginal GWAS in the UKB. Our finding highlights the

importance of accounting for heterogeneity in genetic effect across exposure group in association discovery studies.

The 19q13.2 region contains several genes related to smoking behavior. rs2604894 is an intron variant of *EGLN2. CYP2A6* adjacent to *EGLN2* is involved in nicotine metabolism [25]. A previous study reported that *EGLN2* variants were not associated with nicotine metabolism but with CPD independent of *CYP2A6* variants [26]. Significant lung expression quantitative traits loci (eQTLs) were detected with *EGLN2* [27]. *EGLN2* is known to be involved in regulating hypoxia tolerance and apoptosis in cardiac and skeletal muscle. Further functional studies on 19q13.2 region are clearly warranted to verify the contribution of susceptibility genes in COPD.

We identified genome-wide significant smoking interaction effects at 15q25.1 in UKB and replicated in SpiroMeta, presenting associations primarily in ever-smokers. The *CHRNA5/A3/B4* gene cluster on 15q25.1 encodes the nicotinic acetylcholine receptor subunits α5, α3 and β4. Variants in this gene cluster have been robustly associated with several lung-related traits, such as lung cancer [28] and COPD [6] as well as smoking-related phenotypes, such as smoking quantity [21,29–31] and nicotine dependence [29]. Because smoking is the most important environmental risk factor for COPD, it is quite likely that the association of variants in 15q25.1 region with COPD mediates through smoking behavior [32].

We noted a significant dose response of rs12440014 at 15q25.1 in UKB but not in COPDGene. It may be simply due to a smaller sample size in COPDGene. However, given that COPDGene is enriched for heavy smokers and thus COPD cases compared to UKB, a population-based cohort, our results may suggest other possibilities; 1) the genetic susceptibility of 15q25.1 region

to COPD would be substantial at relatively low level of smoking exposure and/or 2) COPD patients are likely to quit smoking, diluting the association between 15q25.1 and COPD.

We confirmed a known interaction on COPD, *SERPINA1* (PI*Z allele)-by-smoking interaction in our study population [1,2]. In previous study, a PI*Z-by-smoking interaction was identified on FEV_1 (P=0.03) and COPD status (P=0.01) in subjects of European ancestry [2]. The *SERPINA1*, encoding the AAT protein, influences the risk to COPD [33]. The homozygosity for PI*Z allele is the most common cause of AAT deficiency. It follows a Mendelian pattern of inheritance, but there is marked variability in the development and severity of COPD in PI*ZZ individuals. Our replication can help to understand variable manifestations of COPD among individuals with AAT deficiency.

There are several limitations in our study. First, power for detecting interactions and discovering novel genetic risk factors may be limited even in this large sample size [34]. Second, we only included independent subjects of European ancestry. Investigation of more ethnically diverse populations should lead to more robust inferences of gene-by-environment interaction by increasing diversity of not only environmental exposure but also genetic exposure [35]. Third, our use of two smoking measures, ever smoking and current smoking, in genome-wide investigation may have limited the interpretation of our results. Smokers with more severe COPD are more likely to reduce or quit smoking and those without symptoms are more likely to continue smoking and thus be current smokers, described as “healthy smoker effect”. Such phenomenon is highly possible in COPDGene and may also be relevant for UKB as well [36] . Fourth, we used self-reported smoking history. Measurement errors of smoking may lead to our lack of findings of interaction [7]. Fifth, a “healthy volunteer” selection bias exists in UKB study. The UKB cohort is not representative of the general population; UKB participants are less likely to smoke and have

fewer self-reported health conditions compared with the general population [37]. However, generalizability is not necessary to make an inference of associations. Its large sample size and heterogeneity of smoking exposures would still make our findings valid.

Despite our large sample size (n=200,766), we detected only one locus reaching conventional genome-wide significance for the 1df test for gene-by-smoking interaction. To understand gene-by-smoking interaction on COPD risk or more broadly, gene-by-environment interaction, other approaches should also be considered. First, integration of genetic markers and other "-omics" data (transcriptomic, proteomic or epigenomic data) would be helpful. For example, genetic markers influencing other biomarkers such as eQTL may be more likely to interact with smoking [38]. Second, rare variants could also play a significant role on gene-by-environment interaction. Third, to define the sub-population at highest risk of COPD, a polygenic risk score (PRS) approach might be better useful than a genome-wide approach using a single variant with relatively small effect size [39].

In summary, a genome-wide investigation incorporating gene-by-smoking interaction identified two COPD loci, 16q22.1 – *SMPD3* and 19q13.2 – *EGLN2*. We observed the 19q13.2 region in our joint test which did not reach genome-wide significance in previous marginal GWAS in UKB, suggesting the importance of accounting for heterogeneity in genetic effect across exposure group in association discovery studies. We detected a genome-wide significant interaction at the 15q25.1 region primarily having an association among ever-smokers. We replicated one known interaction, PI*Z allele-by-smoking interaction on COPD. Cigarette smoking is the most important environmental risk factor for COPD, but individuals vary in their susceptibility to the effects of cigarette smoke. It raises the possibility of detectable gene-by-smoking interactions, but we identified few significant interactions in our large-scale study. Considering diverse

populations and other approaches such as integration of genetic data and other "-omics" data and PRS may better help to elucidate gene-by-environment interaction on COPD risk.

References

1. Silverman EK, Province MA, Campbell EJ, Pierce JA, Rao DC, Boerwinkle E. Family study of α1-antitrypsin deficiency: Effects of cigarette smoking, measured genotype, and their interaction on pulmonary function and biochemical traits. *Genet Epidemiol*. 1992;9(5):317-331. doi:10.1002/gepi.1370090504

2. Castaldi PJ, Demeo DL, Hersh CP, et al. Impact of non-linear smoking effects on the identification of gene-by-smoking interactions in COPD genetics studies. *Thorax*. 2011;66(10):903-909. doi:10.1136/thx.2010.146118

3. Deng X, Yuan CH, Chang D. Interactions between single nucleotide polymorphism of SERPINA1 gene and smoking in association with COPD: A case–control study. *Int J COPD*. 2017;12:259-265. doi:10.2147/COPD.S116313

4. Hancock DB, Soler Artigas MM, Gharib SA, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet*. 2012;8(12):e1003098. doi:10.1371/journal.pgen.1003098

5. Park B, Koo S-M, An J, et al. Genome-wide assessment of gene-by-smoking interactions in COPD. *Sci Rep*. 2018;8(1):9319. doi:10.1038/s41598-018-27463-5

6. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet*. 2019;51(3):494-505. doi:10.1038/s41588-018-0342-2

7. Aschard H, Lutz S, Maus B, et al. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Hum Genet*. 2012;131(10):1591-1613. doi:10.1007/s00439-012-1192-0

8. Kraft P, Yen Y-CC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007;63(2):111-119. doi:10.1159/000099183

9. Sidney S, de Faire U, Faul JD, et al. Multi-Ancestry Genome-Wide Association Study of Lipid Levels Incorporating Gene-Alcohol Interactions. *Am J Epidemiol*. January 2019. doi:10.1093/aje/kwz005

10. Bentley AR, Sung YJ, Brown MMRM, et al. Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nat Genet*. 2019;51(4):636-648. doi:10.1038/s41588-019-0378-y

11. Xu J, Gaddis NC, Bartz TM, et al. Omega-3 Fatty Acids and Genome-wide Interaction Analyses Reveal DPP10 -Pulmonary Function Association. *Am J Respir Crit Care Med*. 2018;30:rccm.201802-0304OC. doi:10.1164/rccm.201802-0304OC

12. Sung YJ, Winkler TW, de Las Fuentes L, et al. A Large-Scale Multi-ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. *Am J Hum Genet*. 2018;102(3):375-400. doi:10.1016/j.ajhg.2018.01.015

13. Justice AE, Winkler TW, Feitosa MF, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun*. 2017;8:14977. doi:10.1038/ncomms14977

14. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z

15. Regan E a, Hokanson JE, Murphy JR, et al. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD J Chronic Obstr Pulm Dis*. 2011;7(1):32-43. doi:10.3109/15412550903499522

16. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019;51(3):481-493. doi:10.1038/s41588-018-0321-7

17. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med*. 2017;195(5):557-582. doi:10.1164/rccm.201701-0218PP

18. Hankinson JL, Odencrantz JR, Fedan KB. *Spirometric Reference Values from a Sample of the General U.S. Population*. Vol 159.; 1999. www.atsjournals.org. Accessed September 3, 2019.

19. Sung YJ, Winkler TW, Manning AK, et al. An Empirical Comparison of Joint and Stratified Frameworks for Studying G × E Interactions: Systolic Blood Pressure and Smoking in the CHARGE Gene-Lifestyle Interactions Working Group. *Genet Epidemiol*. 2016;40(5):404-415. doi:10.1002/gepi.21978

20. Hancock DB, Reginsson GW, Gaddis NC, et al. Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. *Transl Psychiatry*. 2015;5:e651. doi:10.1038/tp.2015.149

21. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2019;51(2):237-244. doi:10.1038/s41588-018-0307-5

22. Bhatt SP, Kim Y, Harrington KF, et al. Smoking duration alone provides stronger risk estimates of chronic obstructive pulmonary disease than pack-years. *Thorax*. 2018;73(5):414-421. doi:10.1136/thoraxjnl-2017-210722

23. Cho MH, Castaldi PJ, Wan ES, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet*. 2012;21(4):947-957. doi:10.1093/hmg/ddr524

24. Vestbo J, Anderson W, Coxson HO, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J*. 2008;31(4):869-873. doi:10.1183/09031936.00111707

25. Nakajima M, Yamamoto T, Nunoya K, et al. Role of human cytochrome P4502A6 in C-oxidation of nicotine. *Drug Metab Dispos*. 1996;24(11):1212-1217. http://www.ncbi.nlm.nih.gov/pubmed/8937855. Accessed August 10, 2019.

26. Bloom AJ, Baker TB, Chen LS, et al. Variants in two adjacent genes, EGLN2 and CYP2A6, influence smoking behavior related to disease risk via different mechanisms. *Hum Mol Genet*. 2014;23(2):555-561. doi:10.1093/hmg/ddt432

27. Lamontagne M, Couture C, Postma DS, et al. Refining Susceptibility Loci of Chronic Obstructive Pulmonary Disease with Lung eqtls. Miao X-P, ed. *PLoS One*. 2013;8(7):e70220. doi:10.1371/journal.pone.0070220

28. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*. 2008;452(7187):633-

637. doi:10.1038/nature06885

29. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008;452(7187):638-642. doi:10.1038/nature06846

30. Furberg H, Kim Y, Dackor J, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010;42(5):441-447. doi:10.1038/ng.571

31. Lutz SM, Frederiksen B, Begum F, et al. Common and Rare Variants Genetic Association Analysis of Cigarettes per Day Among Ever-Smokers in Chronic Obstructive Pulmonary Disease Cases and Controls. *Nicotine Tob Res*. 2019;21(6):714-722. doi:10.1093/ntr/nty095

32. Lutz SM, Hokanson JE. Genetic influences on smoking and clinical disease: Understanding behavioral and biological pathways with mediation analysis. *Ann Am Thorac Soc*. 2014;11(7):1082-1083. doi:10.1513/AnnalsATS.201407-315ED

33. DeMeo DL, Silverman EK. α1-Antitrypsin deficiency • 2: Genetic aspects of α1-antitrypsin deficiency: Phenotypes and genetic modifiers of emphysema risk. *Thorax*. 2004;59(3):259-264. doi:10.1136/thx.2003.006502

34. Burton PR, Hansell AL, Fortier I, et al. Size matters: Just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol*. 2009;38(1):263-273. doi:10.1093/ije/dyn147

35. Ritz BR, Chatterjee N, Garcia-Closas M, et al. Lessons Learned From Past Gene-Environment Interaction Successes. *Am J Epidemiol*. 2017;186(7):778-786. doi:10.1093/aje/kwx230

36. Becklake MR, Lalloo U. The 'Healthy Smoker': A Phenomenon of Health Selection? *Respiration*. 1990;57(3):137-144. doi:10.1159/000195837

37. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *Am J Epidemiol*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

38. Favé M-J, Lamaze FC, Soave D, et al. Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nat Commun*. 2018;9(1):827.

doi:10.1038/s41467-018-03202-2

39. Aschard H, Tobin MD, Hancock DB, et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int J Epidemiol*. 2017;46(3):894-904. doi:10.1093/ije/dyw318

Table 2.1. Subject characteristics stratified by smoking status

	UK Biobank			COPDGene	
	Never	**Former**	**Current**	**Former**	**Current**
N	129175	57001	14590	3361	1981
Moderate COPD	8631 (6.7)	7857 (13.8)	4589 (31.5)	1831 (54.5)	978 (49.4)
Age	55.55 (8.05)	57.91 (7.64)	54.12 (8.06)	64.95 (8.23)	57.48 (7.80)
Female	50194 (38.9)	29175 (51.2)	6941 (47.6)	1620 (48.2)	907 (45.8)
Pack-years	0.00 (0.00)	19.21 (17.50)	28.40 (18.25)	47.15 (27.30)	48.13 (24.49)
BMI	26.87 (4.56)	27.95 (4.58)	26.73 (4.68)	28.97 (5.89)	27.58 (5.79)
FEV1 % predicted	96.44 (13.95)	93.52 (16.30)	85.10 (18.51)	70.03 (29.35)	75.29 (25.05)
FEV1/FVC	0.77 (0.06)	0.76 (0.07)	0.72 (0.09)	0.61 (0.19)	0.65 (0.16)

Mean (SD) for continuous variable; N (%) for categorical variable; FEV1=Forced expiratory volume in one second; FVC=Forced vital capacity

Table 2.2 Significant results of 2df join test and 1df interaction test in UK Biobank

						Genetic Main		Interaction		2DF Joint
rsID	**Chr:Position**	**Nearest Gene**	**Effect/Ref. Allele**	**EAF**	**Smoking Exposure**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**	**P**
Significant from 2df joint test*										
rs141322661	16:68398875	*SMPD3*	G/A	0.01	Ever Smoking	0.76 (0.65-0.88)	2.09E-04	0.94 (0.78-1.15)	5.72E-01	3.92E-09
rs141322661	16:68398875	*SMPD3*	G/A	0.01	Current Smoking	0.77 (0.69-0.86)	1.47E-06	0.81 (0.62-1.07)	1.45E-01	1.45E-08
rs2604894	19:41292404	*EGLN2*	A/G	0.45	Current Smoking	0.96 (0.94-0.98)	1.28E-03	0.9 (0.84-0.95)	4.11E-04	5.87E-09
Significant from 1df interaction test*										
rs7170068	15:78912943	*CHRNA3*	A/G	0.22	Current Smoking	0.96 (0.93-0.98)	2.38E-03	0.79 (0.74-0.86)	1.82E-09	6.08E-16
rs12440014	15:78926726	*CHRNB4*	G/C	0.24	Ever Smoking	1.02 (0.98-1.06)	3.14E-01	0.83 (0.79-0.88)	8.96E-12	6.96E-18
Significant from 1df interaction test in candidate variants**										
rs7642001	3:168746145	*MECOM*	A/G	0.37	Current Smoking	1.07 (1.04-1.09)	3.20E-07	1.12 (1.05-1.19)	3.65E-04	1.73E-14
rs28929474	14:94844947	*SERPINA1*	T/C	0.02	Ever Smoking	0.95 (0.85-1.07)	4.02E-01	1.3 (1.12-1.52)	6.70E-04	1.10E-04

EAF: Effect Allele Frequency; Interaction: Interaction test with 1 degree of freedom; 2DF Joint: Joint test with 2 degrees of freedom of genetic main and interaction effects; *Genome-wide statistical significance (P<5.00E-08) applied; **Bonferroni corrected statistical significance (P<1.00E-04) applied.

Table 2.3. Replications of selected SNPs in COPDGene and SpiroMeta

						COPDGene NHW		SpiroMeta (FEV1)					
						Marginal Association		Never Smoker		Ever Smoker		Interaction	2DF Joint
rsID	Chr:Position	Nearest Gene	Effect/ Ref. Allele	EAF	Smoking Exposure	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	P	P
Main effects													
rs141322661	16:68398875	*SMPD3*	G/A	0.01	Ever Smoking	0.99 (0.69-1.41)	9.52E-01	0.01 (-0.05, 0.08)	6.45E-01	0.06 (0, 0.13)	6.03E-02	1.07E-01	1.54E-01
rs2604894	19:41292404	*EGLN2*	A/G	0.45	Current Smoking	0.9 (0.82-0.98)	1.62E-02	0.001 (-0.02, 0.02)	9.12E-01	0.003 (-0.01, 0.02)	6.91E-01	7.66E-01	9.18E-01
Evidence with interaction effects													
rs7642001	3:168746145	*MECOM*	A/G	0.37	Current Smoking	1.1 (1.01-1.2)	3.63E-02	-0.04 (-0.05, -0.02)	3.31E-05	-0.01 (-0.03, 0.01)	1.93E-01	1.07E-03	7.79E-05
rs28929474	14:94844947	*SERPINA1*	T/C	0.02	Ever Smoking	1.34 (1-1.81)	5.08E-02	0.04 (-0.02, 0.09)	2.04E-01	0.02 (-0.04, 0.08)	4.65E-01	5.53E-01	3.41E-01
rs12440014	15:78926726	*CHRNB4*	G/C	0.24	Ever Smoking	0.76 (0.69-0.85)	3.41E-07	0.01 (-0.01, 0.03)	4.95E-01	0.03 (0.01, 0.05)	1.46E-03	7.33E-03	5.00E-03

EAF: Effect Allele Frequency; Interaction: Interaction test with 1 degree of freedom; 2DF Joint: Joint test with 2 degrees of freedom of genetic main and interaction effects; Marginal association between each selected SNP and COPD was tested in COPDGene; Lookup of selected SNPs in GWAS of spirometric measures of lung function stratified by never- and ever-smoker groups in SpiroMeta; EAFs from COPDGene and SpiroMeta were similar.

Figure 2.1. Analysis workflow

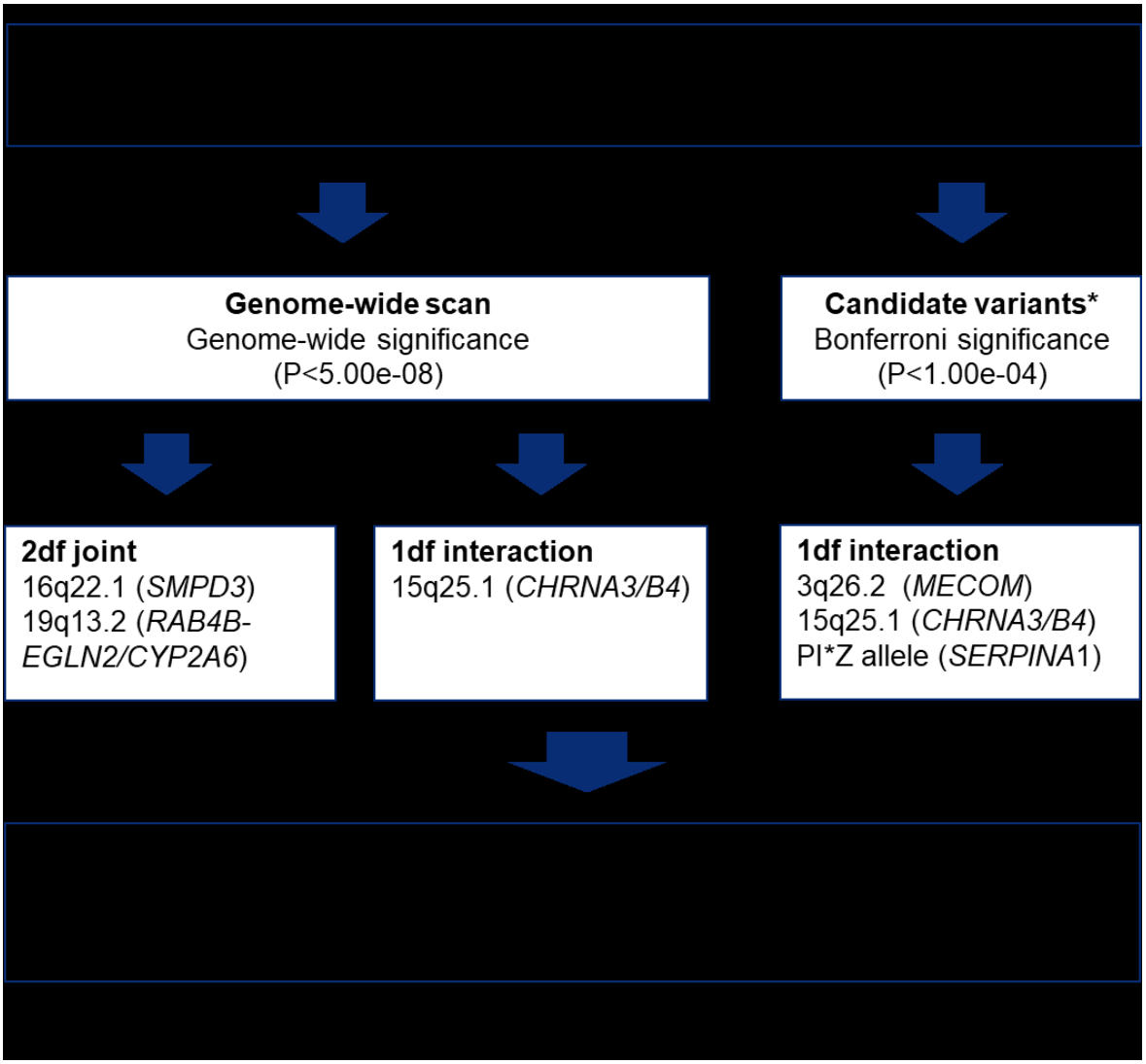


Figure 2.2. Regional plots of 16q22.1 and 19q13.2 region based on 2df joint test

Figure 2.2a. 16q22.1 region

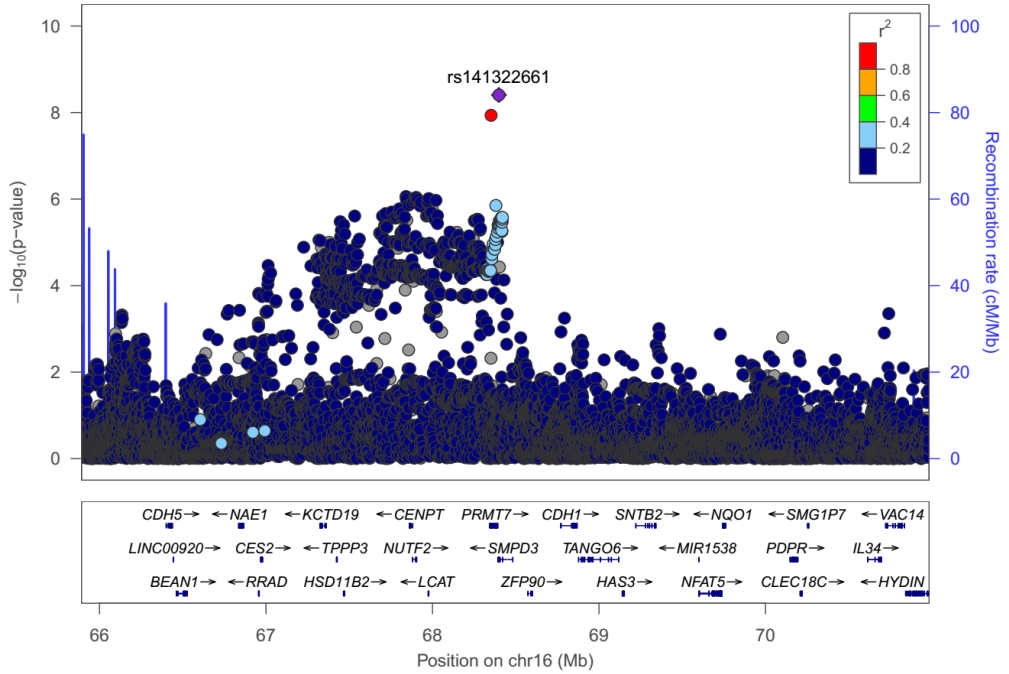


Figure 2.2b. 19q13.2 region

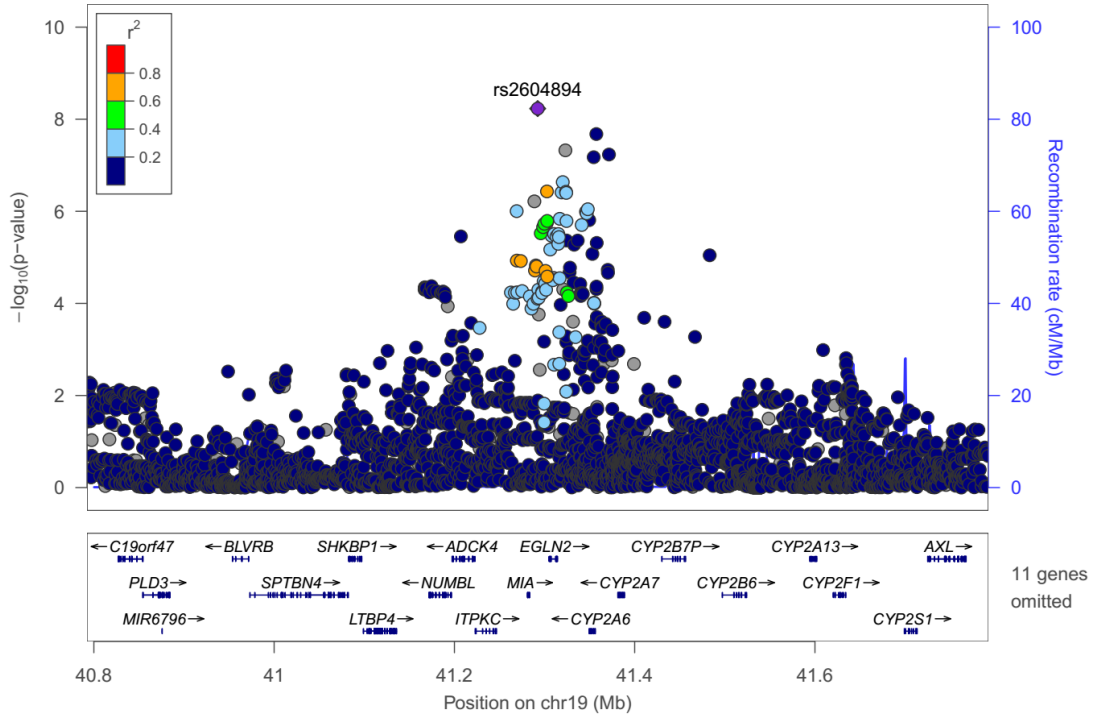


Figure 2.3. Manhattan plot and regional plot of 15q25 region based on 1df interaction test

Figure 2.3a. Manhattan plot of 15q25 region

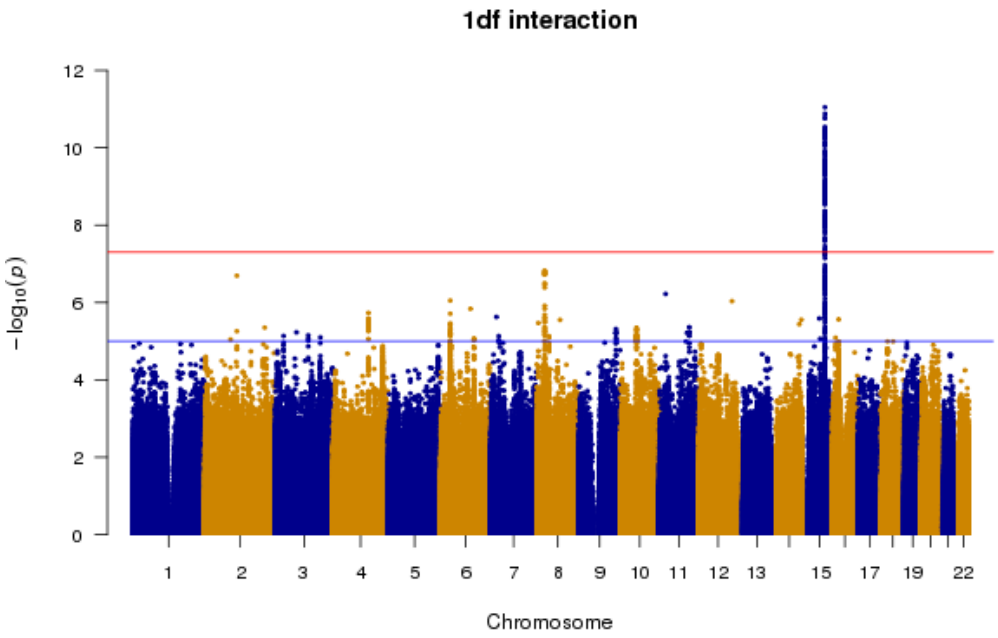


Figure 2.3b. Regional plot of 15q25 region

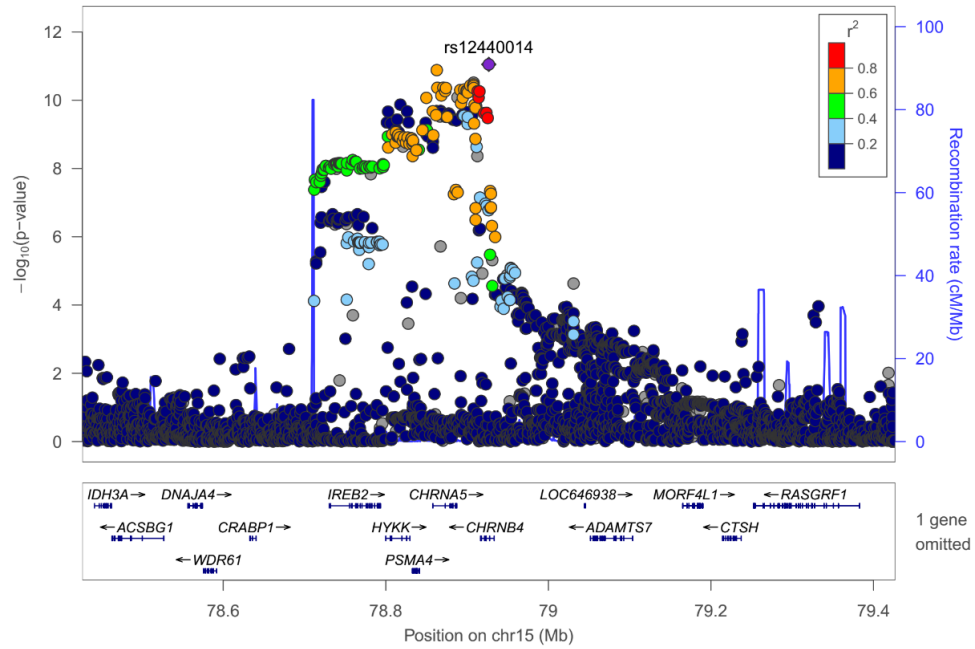
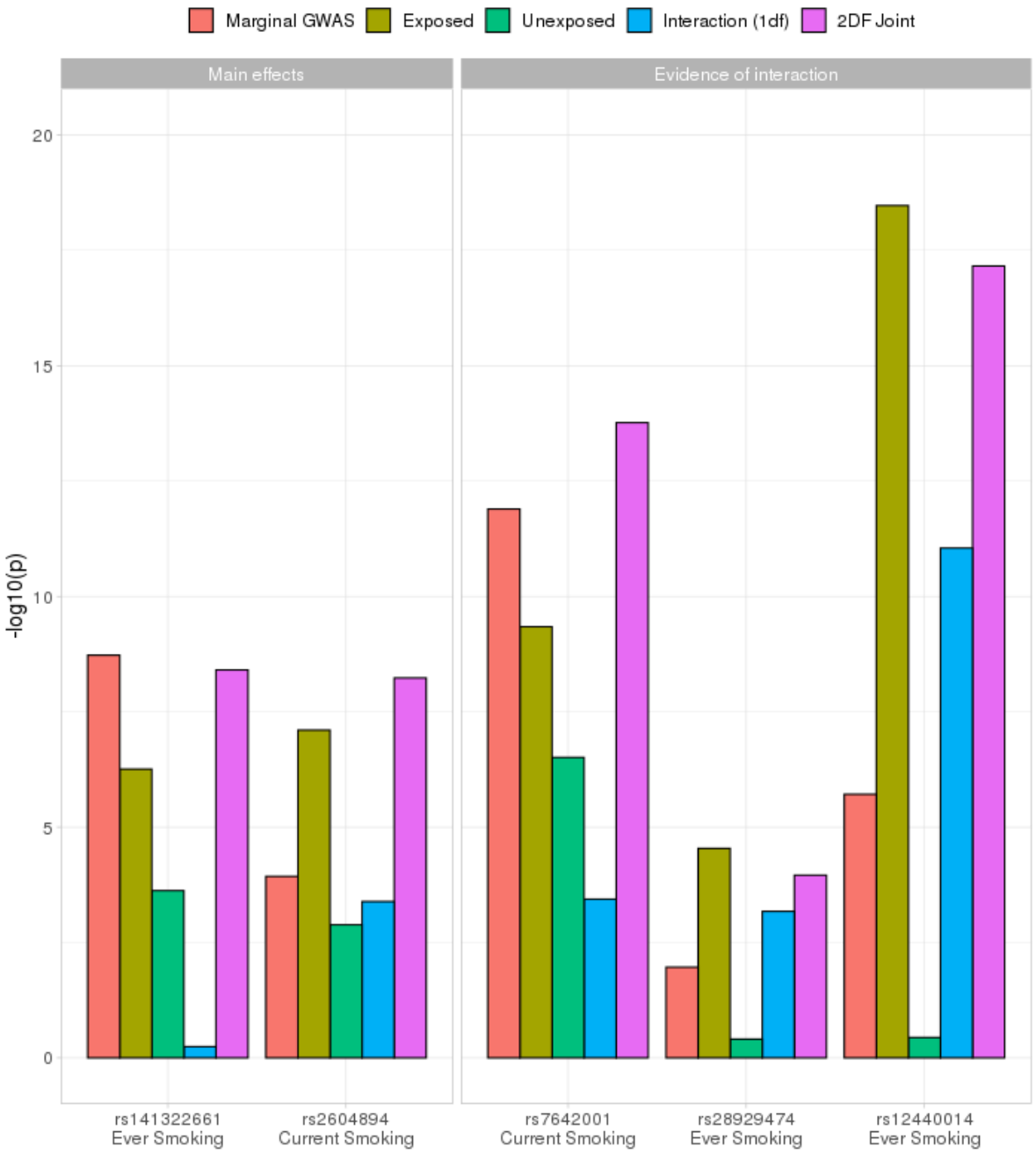


Figure 2.4. Statistical significances of selected SNPs



Supplemental Table 2.1. Lookup in GWAS of Smoking (Liu et al., 2019)

Trait	Variant	Chromosome	Position	Effect Allele	EAF	Beta	P
cigarettes per day	rs12438181	15	78812098	A	0.22	-0.01853	4.97E-10
cigarettes per day	rs10519203	15	78814046	A	0.66	-0.09362	3.12E-286
cigarettes per day	rs28438420	15	78836288	T	0.55	0.01756	1.25E-12
cigarettes per day	rs72740955	15	78849779	T	0.34	0.03175	2.42E-34
cigarettes per day	rs146009840	15	78906177	T	0.34	0.02212	2.00E-17
cigarettes per day	rs28681284	15	78908565	T	0.21	-0.04868	2.10E-58
cigarettes per day	rs8040868	15	78911181	C	0.4	0.01601	1.79E-10
cigarettes per day	rs3743063	15	79065171	C	0.56	-0.01672	1.53E-11
cigarettes per day	rs143200968	19	41338847	C	0.03	-0.0861	6.97E-28
cigarettes per day	rs56113850	19	41353107	C	0.56	0.05231	4.01E-99
cigarettes per day	rs8192726	19	41354496	A	0.07	-0.03934	8.35E-16
cigarettes per day	rs117824460	19	41371480	G	0.03	-0.09526	7.66E-35
current vs. former	rs518425	15	78883813	G	0.28	-0.0305	1.72E-12
current vs. former	rs145580088	19	41342842	G	0.02	0.09096	9.48E-13
current vs. former	rs59586387	19	41375030	G	0.07	0.05139	3.37E-11

EAF: Effect Allele Frequency

Supplemental Table 2.2. Interactions of selected SNPs with smoking dose measures in UK Biobank and COPDGene

						UK Biobank (n=200,766)				UK Biobank ever smokers (n=67,761)				COPDGene smokers (n=5,342)			
						Quantitative		Ordinal**		Quantitative		Ordinal***		Quantitative		Ordinal***	
rsID	Chr.	Position	Nearest Gene	Risk Allele*	Ref. Allele	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P
Smoking Duration																	
rs7642001	3	168746145	MECOM	A	G	1.0015 (1.0001-1.0028)	**3.00E-02**	1.0138 (1-1.0277)	**4.92E-02**	1.0035 (1.0008-1.0062)	**1.01E-02**	1.0413 (1.0092-1.0744)	**1.12E-02**	1.0164 (1.007-1.026)	**6.20E-04**	1.1212 (1.0327-1.2173)	**6.42E-03**
rs28929474	14	94844947	SERPINA1	T	C	1.0083 (1.0037-1.013)	**3.89E-04**	1.0814 (1.0312-1.1341)	**1.25E-03**	1.0005 (0.9916-1.0094)	9.16E-01	0.9567 (0.8643-1.0589)	3.92E-01	1 (0.9701-1.0307)	9.98E-01	1.1548 (0.8592-1.552)	3.40E-01
rs12440014	15	78926726	CHRNB4	C	G	1.0057 (1.0041-1.0072)	**5.42E-13**	1.064 (1.0472-1.0812)	**2.63E-14**	1.0041 (1.001-1.0072)	**1.06E-02**	1.0633 (1.0248-1.1032)	**1.12E-03**	1.0081 (0.9973-1.0191)	1.43E-01	1.0415 (0.9455-1.1472)	4.10E-01
rs141322661	16	68398875	SMPD3	A	G	1.0009 (0.9951-1.0067)	7.57E-01	1.0041 (0.9455-1.0664)	8.93E-01	1.0026 (0.991-1.0144)	6.57E-01	0.9988 (0.8684-1.1487)	9.86E-01	0.9626 (0.92-1.0071)	9.82E-02	0.7678 (0.5356-1.1008)	1.51E-01
rs2604894	19	41292404	EGLN2	G	A	1.0017 (1.0004-1.003)	**1.12E-02**	1.0158 (1.0022-1.0296)	**2.24E-02**	0.9999 (0.9973-1.0025)	9.48E-01	0.9857 (0.9558-1.0166)	3.61E-01	0.997 (0.988-1.006)	5.12E-01	0.9718 (0.8972-1.0525)	4.82E-01
Pack-years																	
rs7642001	3	168746145	MECOM	A	G	1.001 (0.9999-1.0022)	8.12E-02	1.0103 (0.9967-1.0241)	1.38E-01	1.0011 (0.9994-1.0027)	2.10E-01	1.0223 (0.9916-1.054)	1.57E-01	1.0069 (1.0025-1.0113)	**2.11E-03**	1.1085 (1.0207-1.2039)	**1.44E-02**
rs28929474	14	94844947	SERPINA1	T	C	1.0058 (1.0017-1.0099)	**5.42E-03**	1.0786 (1.0294-1.1302)	**1.50E-03**	1.0001 (0.9944-1.0058)	9.83E-01	0.9758 (0.883-1.0783)	6.31E-01	0.9989 (0.9838-1.0143)	8.91E-01	0.9156 (0.6991-1.1993)	5.22E-01
rs12440014	15	78926726	CHRNB4	C	G	1.0034 (1.002-1.0048)	**8.32E-07**	1.0506 (1.0339-1.0674)	**1.34E-09**	1.0012 (0.9993-1.0032)	2.23E-01	1.0228 (0.9865-1.0604)	2.21E-01	1.0022 (0.9971-1.0073)	3.99E-01	1.0571 (0.9595-1.1646)	2.61E-01
rs141322661	16	68398875	SMPD3	A	G	0.9988 (0.9938-1.0038)	6.31E-01	0.9987 (0.9412-1.0596)	9.65E-01	0.9975 (0.9902-1.0048)	5.00E-01	0.9817 (0.8577-1.1236)	7.89E-01	0.9927 (0.9726-1.0132)	4.81E-01	0.8367 (0.5826-1.2016)	3.34E-01
rs2604894	19	41292404	EGLN2	G	A	1.0006 (0.9995-1.0017)	2.88E-01	1.0136 (1.0002-1.0272)	**4.73E-02**	0.9997 (0.9981-1.0013)	7.31E-01	0.999 (0.9695-1.0294)	9.49E-01	1.0003 (0.9961-1.0046)	8.83E-01	0.9946 (0.9172-1.0785)	8.95E-01
Cigarette Per Day																	
rs7642001	3	168746145	MECOM	A	G	1.0006 (0.9989-1.0024)	5.00E-01	1.0029 (0.9891-1.017)	6.81E-01	0.9993 (0.9964-1.0023)	6.54E-01	0.9877 (0.9608-1.0153)	3.78E-01	1.0035 (0.9957-1.0115)	3.78E-01	1.0294 (0.9552-1.1094)	4.48E-01
rs28929474	14	94844947	SERPINA1	T	C	1.0074 (1.0015-1.0134)	**1.47E-02**	1.0826 (1.0325-1.1351)	**1.03E-03**	0.9961 (0.9863-1.006)	4.41E-01	1.0089 (0.9186-1.1081)	8.53E-01	0.9888 (0.9626-1.0157)	4.11E-01	1.0038 (0.7749-1.3004)	9.77E-01
rs12440014	15	78926726	CHRNB4	C	G	1.0048 (1.0027-1.0069)	**5.91E-06**	1.0383 (1.0213-1.0556)	**8.34E-06**	1.0007 (0.9972-1.0043)	6.95E-01	0.9854 (0.9535-1.0185)	3.84E-01	1.0006 (0.9914-1.0099)	9.00E-01	1.0081 (0.9216-1.1027)	8.60E-01
rs141322661	16	68398875	SMPD3	A	G	1 (0.9926-1.0076)	9.91E-01	0.9832 (0.9251-1.045)	5.87E-01	0.9971 (0.9847-1.0096)	6.45E-01	0.9234 (0.8172-1.0435)	2.02E-01	1.0106 (0.9781-1.0442)	5.27E-01	0.9884 (0.706-1.3839)	9.46E-01
rs2604894	19	41292404	EGLN2	G	A	1.0005 (0.9988-1.0022)	5.47E-01	1.0117 (0.9979-1.0257)	**9.62E-02**	0.9986 (0.9957-1.0015)	3.39E-01	0.9975 (0.9708-1.0251)	8.59E-01	1.0025 (0.9946-1.0106)	5.31E-01	1.0235 (0.95-1.1027)	5.41E-01
Fagerstrom Test for Nicotine Dependence (FTND) score																	
rs7642001	3	168746145	MECOM	A	G									0.9489 (0.8961-1.0048)	7.27E-02	0.8441 (0.7017-1.0155)	7.24E-02
rs28929474	14	94844947	SERPINA1	T	C									0.9146 (0.7099-1.1784)	4.90E-01	0.7475 (0.325-1.7189)	4.93E-01
rs12440014	15	78926726	CHRNB4	C	G									1.0733 (1.002-1.1496)	**4.37E-02**	1.1607 (0.9316-1.4463)	1.84E-01
rs141322661	16	68398875	SMPD3	A	G									0.9484 (0.76-1.1835)	6.39E-01	0.7671 (0.3747-1.5705)	4.68E-01
rs2604894	19	41292404	EGLN2	G	A									1.0101 (0.9552-1.0682)	7.25E-01	0.9794 (0.8151-1.1767)	8.24E-01

Smoking Duration, Pack-years and Cigarette per day were categorized into quartiles (Q1, Q2, Q3 and Q4).; FTND score was categorized into mild (0-3), moderate (4-6) or severe (7-10).; *Set the allele increasing the risk; **Never smoker group was set as a reference group; ***The lowest group of smoking dose (Q1) was set as a reference group; Ref. Allele: Reference Allele; 1937 COPDGene subjects were included for FTND score analysis.

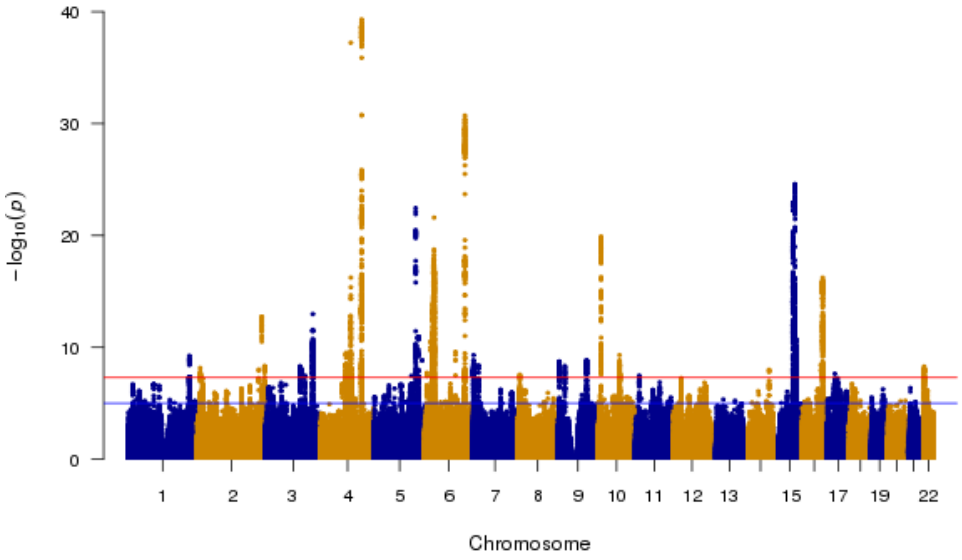
Supplemental Table 2.3. Stratified analysis by ever smoking for selected SNPs in SpiroMeta (FVC and FEV_1/FVC)

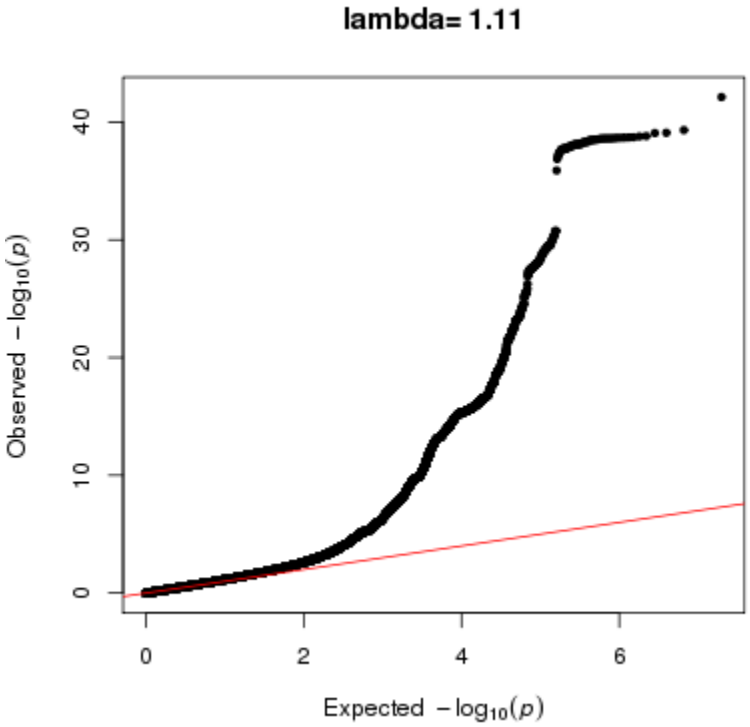
			Never Smoker		Ever Smoker		1DF Interaction	2DF Joint
rsID	**Chr:Position**	**Effect/Ref. Allele**	**Beta (95% CI)**	**P**	**Beta (95% CI)**	**P**	**P**	**P**
FVC								
rs7642001	3:168746145	A/G	-0.0242 (-0.0408, -0.0075)	4.54E-03	-0.0039 (-0.0201, 0.0122)	6.35E-01	3.90E-02	1.59E-02
rs11735046	4:145008115	A/G	0.0079 (-0.0188, 0.0346)	5.61E-01	-0.0236 (-0.05, 0.0028)	8.02E-02	4.72E-02	1.83E-01
rs56326741	6:32067967	T/C	0.0145 (-0.0128, 0.0418)	2.99E-01	0.0044 (-0.023, 0.0318)	7.55E-01	5.37E-01	5.56E-01
rs55676755	15:78898932	G/C	-0.0077 (-0.0248, 0.0093)	3.74E-01	-0.0293 (-0.0459, -0.0126)	5.76E-04	3.24E-02	1.80E-03
rs28534575	15:78923845	G/T	0.0061 (-0.0134, 0.0257)	5.38E-01	0.0119 (-0.0071, 0.0309)	2.18E-01	6.16E-01	3.88E-01
rs2604894	19:41292404	A/G	-0.002 (-0.018, 0.014)	8.03E-01	-0.006 (-0.0211, 0.009)	4.32E-01	6.66E-01	7.12E-01
FEV_1/FVC								
rs7642001	3:168746145	A/G	-0.0177 (-0.0344, -0.001)	3.74E-02	-0.0119 (-0.0279, 0.0041)	1.45E-01	5.50E-01	3.97E-02
rs11735046	4:145008115	A/G	-0.0258 (-0.0525, 9e-04)	5.82E-02	0.007 (-0.0192, 0.0331)	6.02E-01	3.79E-02	1.45E-01
rs56326741	6:32067967	T/C	0.0182 (-0.0091, 0.0455)	1.92E-01	0.0164 (-0.0108, 0.0435)	2.37E-01	9.11E-01	2.12E-01
rs55676755	15:78898932	G/C	0.0104 (-0.0067, 0.0275)	2.32E-01	-0.0273 (-0.0438, -0.0108)	1.19E-03	1.71E-04	2.57E-03
rs28534575	15:78923845	G/T	0.0066 (-0.013, 0.0261)	5.10E-01	0.0234 (0.0046, 0.0422)	1.46E-02	1.41E-01	4.07E-02
rs2604894	19:41292404	A/G	0.0106 (-0.0054, 0.0267)	1.93E-01	0.0109 (-0.0041, 0.0259)	1.54E-01	9.78E-01	1.55E-01

Ref. Allele: Reference Allele

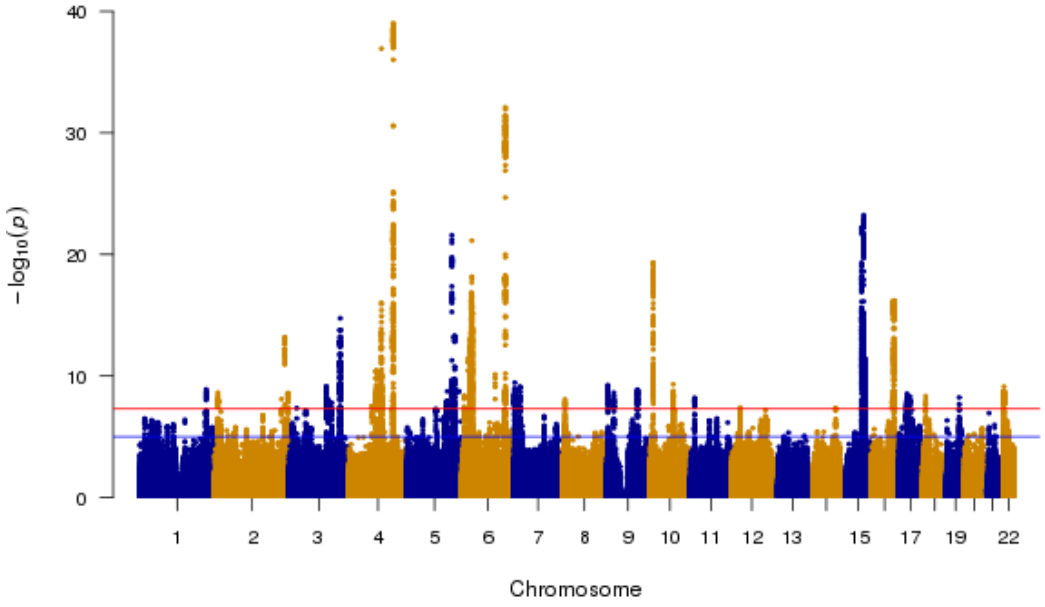
Supplemental Figure 2.1. Manhattan plot and QQ plot of 2df joint test

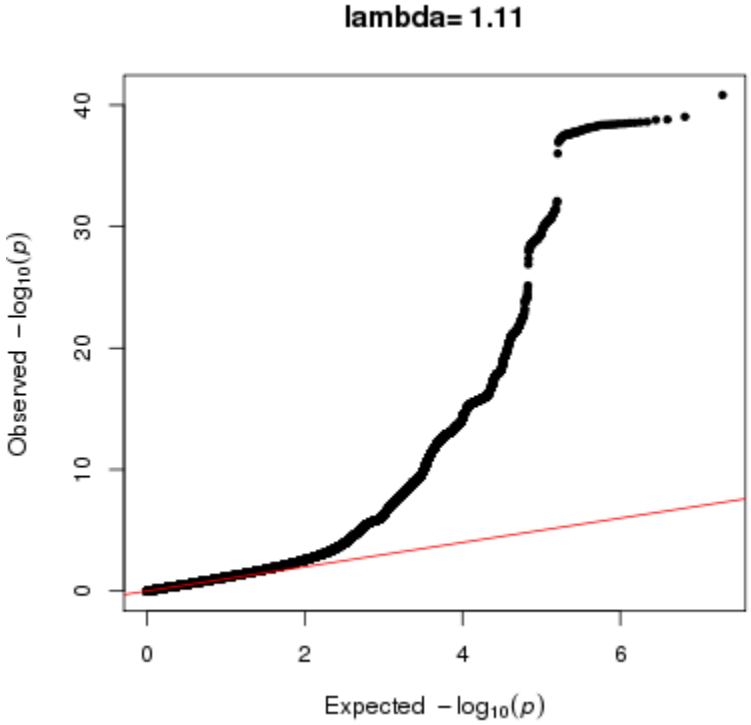
Supplemental Figure 2.1a. Gene-by-ever smoking interaction analysis





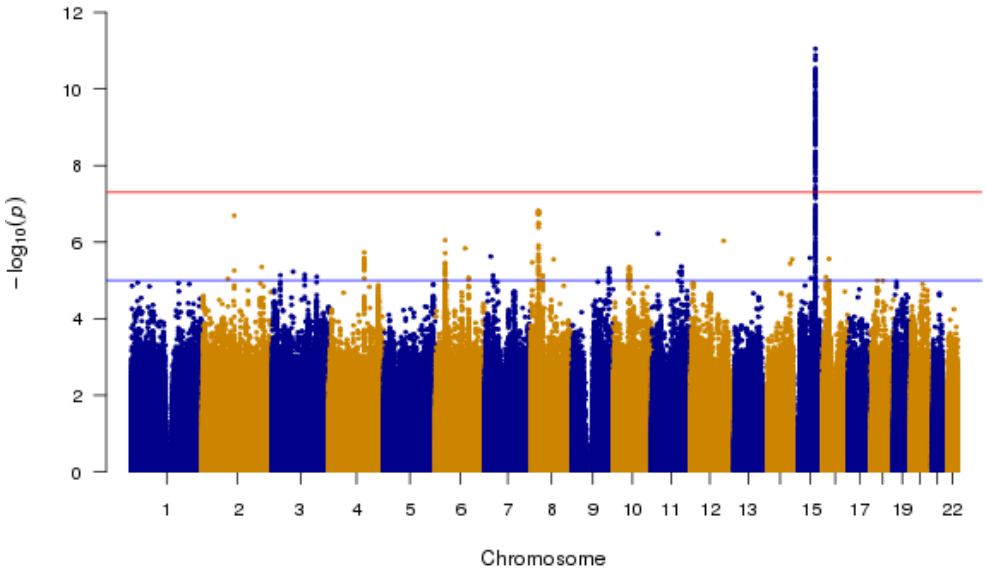
Supplemental Figure 2.1b. Gene-by-current smoking interaction analysis

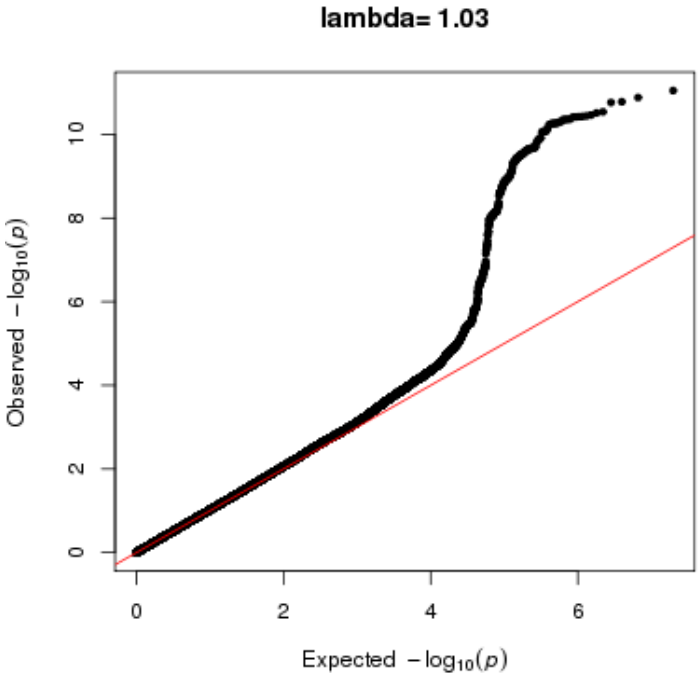




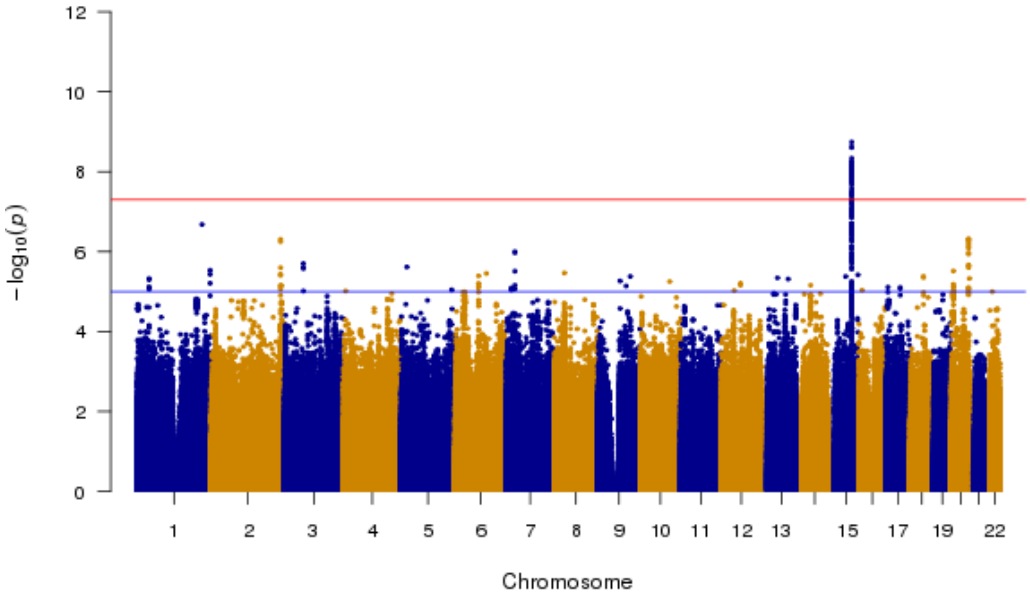
Supplementary Figure 2.2. Manhattan plot and QQ plot of 1df interaction test

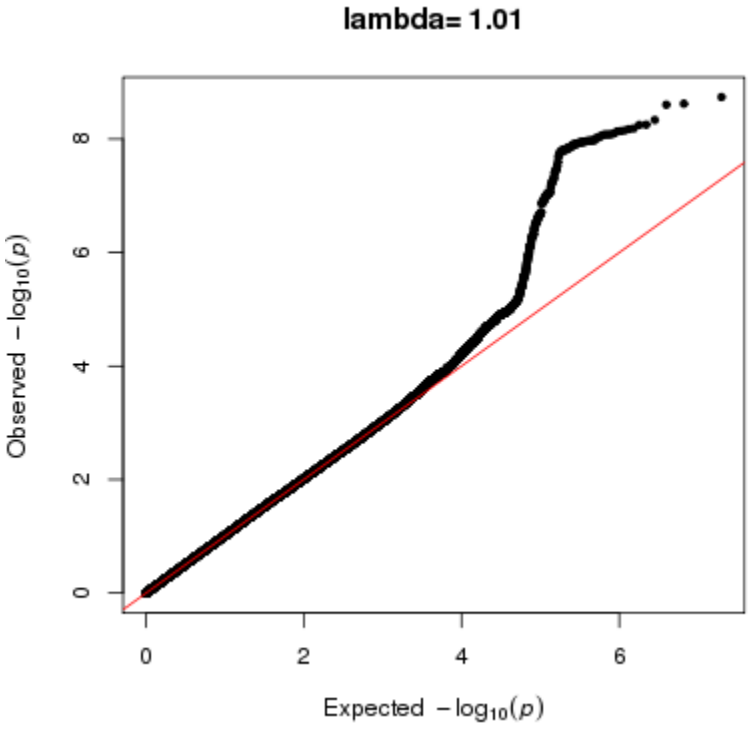
Supplemental Figure 2.2a. Gene-by-ever smoking interaction analysis





Supplemental Figure 2.2b. Gene-by-current smoking interaction analysis





Chapter 4. Gene-by-smoking interaction study on chronic obstructive pulmonary disease using a polygenic risk score of lung function

Chapter 4. Gene-by-smoking interaction study on chronic obstructive pulmonary disease using a polygenic risk score of lung function

Introduction

Complex diseases are influenced by the interplay between environmental and genetic factors. Gene-by-environment interaction is an important component to understand disease mechanism. Statistical interaction can be assessed on multiplicative or additive scales. Additive interactions are less explored than multiplicative interactions in gene-by-environment interaction studies, but such interaction holds greater public health relevance. Additive interaction can be useful when targeting sub-populations where the intervention will be most effective, particularly when resources are constrained [1].

A relatively small effect of a single genetic marker such as single-nucleotide polymorphism (SNP) on an outcome makes it challenging to detect and replicate gene-by-environment interactions. Several studies have attempted to utilize a polygenic risk score (PRS) aggregating genetic variants into a single measure to explore gene-by-environment interactions [2–6]. The PRS is the sum of the number of risk alleles for each SNP identified from genome-wide association study (GWAS). To define the sub-populations more susceptible to a complex disease, this PRS could be a better measure to study gene-by-environment interactions than any single marker.

Risk of Chronic Obstructive Pulmonary Disease (COPD) is influenced by cigarette smoking and genetic susceptibility. The single most important environmental risk factor for COPD is cigarette smoking, but smoking effects vary among individuals possibly due to their genetic makeup. It raises the potential of gene-by-smoking interactions. However, to date little is known about

gene-by-smoking interaction on COPD. A recent study of genome-wide gene-by-smoking interaction on risk to COPD reported evidence of interaction with two genetic regions known to influence risk to COPD, *CHRNB4* and *SERPINA1* in European populations [7]. Other known COPD or lung function risk loci did not seem to interact with smoking on COPD risk.

Aschard et al. 2017 tested for interaction between a PRS of 26 SNPs in relation to forced expiratory volume in 1 second (FEV_1) or the ratio of FEV_1/ forced vital capacity (FVC) identified from a GWAS of lung function [8] and smoking exposure on lung function [5]. A recent large-scale GWAS identified 279 distinct loci associated with lung function and tested for interaction between a PRS based on these 279 lung function SNPs and ever-smoking on quantitative measures of lung function (FEV_1/FVC) and risk to COPD [9].

Expanding previous studies of PRS-by-smoking interaction on lung function and/or COPD, our study tests for potential interactions between a lung function PRS and smoking on risk to COPD under multiplicative and additive models in a large cohort of subjects from the UK Biobank study. Our aim is to examine whether effects of smoking on COPD risk are modified by a lung function PRS. We hypothesize that the increased risk of COPD by smoking would be greater among individuals who are more genetically predisposed.

Methods

Study population

We used data from the UK Biobank (UKB) study, a population-based cohort of volunteers where over 500,000 individuals were originally recruited [10]. Details of quality controls (QC) of spirometric measures, genetic markers and subjects in the UKB study are previously described

[9–11]. Briefly, to determine lung function, measures of FEV_1 and FVC were derived from the spirometry volume-time series data, subjected to additional quality control based on ATS/ERS criteria [9,12]. Genotyping was performed using Axiom UK BiLEVE array and Axiom Biobank array (Affymetrix, Santa Clara, California, USA) and imputed to the Haplotype Reference Consortium (version 1.1) panel. We included independent subjects of European ancestry based on the combination of self-reported ethnicity data and principal components (PCs) data provided by UKB.

Smoking exposures

We assigned smoking status to individuals in UKB based on their responses on questionnaires. Never-smokers included non-current-smokers or those who smoked less than 100 cigarettes in their life. Ever-smokers were defined as either current, most days (current or all days in the past) or smoked occasionally. We defined pack-years as packs of cigarettes smoked per day multiplied by years smoked. Pack-years was 0 for never-smokers. For ever-smokers, we set pack-years to cigarettes per day divided by 20 multiplied by duration of smoking according to individual's smoking status (former or current). We considered never-/ever-smoking status as a binary variable and pack-years as both a quantitative variable and a categorical variable grouped based on ≤10, 10.1-20, 20.1-30, 30.1-40, 40.1-50 and >50 with pack-years, where the lowest exposure group (packyears ≤10) served as the reference group. As a secondary analysis, we considered never-/former-/current-smoking status and pack-years as a categorical variable with pack-years=0 (i.e. never-smoker) as the reference group.

Polygenic risk score based on lung function

Using the 279 SNPs previously associated with lung function in the UKB (where most individuals were identified as having European ancestry), we generated a PRS by summing the

number of risk alleles for each SNP weighted by their estimated effect size for FEV_1/FVC, as previously described [9], in our study population. We divided individuals into deciles according to their value of the PRS. We analyzed PRS as both a quantitative variable and a categorical variable (10 decile groups). Subsequently, we will refer to this lung function PRS as simply PRS.

Outcome

We defined COPD cases based on pre-bronchodilator spirometry following the modified Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria for moderate airflow limitation: FEV_1 less than 80% of predicted value (using reference equations from [13]), and the ratio of FEV_1/FVC less than 0.7.

Statistical analysis

We performed a logistic regression analysis, adjusting for age, age^2, sex, height, genotyping array and the first 10 PCs. We performed two stratified analyses: 1) by PRS group to estimate the odds ratio (OR) of being a COPD case and its 95% confidence interval (95% CI) for the smoking variable, which is relevant to our hypothesis that PRS might be an effect modifier of the association between cigarette smoking and risk to COPD, and 2) by smoking group to estimate OR and 95% CI for the PRS variable, which provides a different perspective that the association between PRS and COPD would differ by smoking behavior.

Tests for interactions. To statistically test for PRS-by-smoking interaction, we included the interaction term (PRS*Smoking) in the model and assessed the multiplicative interaction ($OR_{PRS*Smoking}$) based on maximum likelihood estimation. If the estimated $OR_{PRS*Smoking}>1$, the model would indicate a super-multiplicative interaction, while the estimated $OR_{PRS*Smoking}<1$, a

sub-multiplicative interaction. We tested the additive interaction calculating the Relative Excess Risk due to Interaction (RERI) with the following equations [1].

$$RERI_{OR} = OR_{PRS*Smoking} - OR_{PRS} - OR_{Smoking} + 1$$

We assume the outcome is rare and thus ORs approximate risk ratios (RRs). When the RERI>0, the model would indicate a super-additive interaction, while the RERI<0, a sub-additive interaction. When putting into a public health context, RERI>0 implies the public health consequences of an intervention on smoking would be larger in individuals with a higher genetic risk compared to individuals with a lower genetic risk. The 95% CI of RERI was obtained using the delta method with "nlcom" command in STATA software (version 14.2 https://www.stata.com/) [14].

Joint effects of smoking and PRS. We also examined joint effect of smoking and PRS to assess deviation of the observed joint effect from the expected effect under a multiplicative or additive model as another way to test for statistical interaction. To avoid the high-dimensional issue that would result if all 10 deciles of PRS were modeled, we focused on the two extremes of low and high risk (i.e. 1st and 10th deciles) of PRS. We created a categorical variable with mutually exclusive strata formed by the cross classification of smoking and PRS (10th vs. 1st decile). In this analysis, the reference category becomes the group with lowest smoking exposure (never-smoker or pack-years≤10) in the lowest PRS group. We then regressed COPD case on this categorical variable. The expected OR was calculated as the product of the observed main effects for smoking and PRS under the multiplicative model and as the sum of the main effects minus 1 under the additive model [15]. If the estimated OR is greater than that expected under the multiplicative or additive model, respectively, it would indicate a super-multiplicative or super-additive interaction. If the estimated OR is less, it would indicate a sub-multiplicative or sub-additive interaction.

Population attributable risk. The population attributable risk (PAR) can be interpreted as the proportion of COPD cases expected to occur in a population that is attributable to a particular exposure or combination of exposures. We estimated PARs for smoking, PRS and the combination of both factors. The combined PAR for smoking and PRS was computed from estimates of the main and joint ORs under a logistic regression model using the following equations.

$$PAR = 1 - \sum \frac{\rho_j}{OR_j}$$

For j mutually exclusive strata formed by the cross classification of smoking and PRS, ρ_j is the proportion of all cases in stratum j and OR_j is the OR in group j compared with the reference group [16]. Analyses were performed using R software (version 3.6.0, https://www.r-project.org/) and STATA software.

Results

Basic Characteristics

We used data on 200,766 subjects of European ancestry including 179,689 controls and 21,077 COPD cases from the UKB study (**Table 3.1**). The overall prevalence of COPD was 10.5% in our study population. As expected, the proportion of ever-smokers among controls was lower (32.9%) than among cases (59.1%). The mean pack-years (Mean (SD)) was 6.09 (12.53) in controls and 19.71 (23.91) in COPD cases. The proportion of COPD cases was significantly higher in the 10th decile of the PRS distribution (17.8%) compared to the 1st decile (5.0%).

Overall associations of smoking and lung function PRS with COPD

The estimated OR (95% CI) for risk of COPD comparing ever-smokers to never-smokers was 2.37 (2.30-2.44) (**Table 3.2**). The estimated OR (95% CI) for COPD comparing smokers with very heavy exposure (i.e. pack-years>50) to the reference group of those with less than 10 pack-years was 5.77 (5.27-6.31) among ever-smokers. When treated as a continuous variable, PRS was significantly associated with an increased risk of COPD (OR (95% CI)=1.04 (1.03-1.04)). When PRS was categorized into deciles, the estimated ORs for COPD tended to increase across deciles of PRS; the estimated OR (95% CI) for COPD comparing the 2nd to 10th decile to the lowest group (i.e. 1st decile) was 1.38 (1.27-1.50), 1.52 (1.4-1.66), 1.73 (1.59-1.88), 1.85 (1.71-2.01), 2.20 (2.03-2.38), 2.42 (2.23-2.61), 2.62 (2.42-2.84), 3.11 (2.88-3.36) and 4.35 (4.03-4.68), respectively. Associations of other smoking measures (never/former/current-smoking status and a categorical variable of pack-years with never-smoker as a reference) are provided in **Supplemental Table 3.1**.

Interaction between ever-smoking and PRS

The strength of association between ever-smoking and COPD was greater in the lower decile of PRS (**Figure 3.1a**). Viewed differently, the strength of association between each decile of PRS and risk of COPD was greater among never-smokers compared to ever-smokers across the 2nd to 10th deciles of PRS (**Figure 3.1b**).

In **Table 3.3**, focusing on two extreme groups of PRS (i.e. 1st and 10th decile), the estimated OR (95% CI) for COPD comparing ever-smokers to never-smokers was greater in the lowest PRS group (2.81 (2.46-3.22)) than in the highest PRS group (2.19 (2.02-2.37)). We observed a statistically significant sub-multiplicative interaction (OR (95% CI)=0.79 (0.68-0.92), P=3.11E-03) and super-additive interaction (RERI (95% CI)=4.35 (3.54-5.16), P=6.28E-26). The

estimated OR (95% CI) for joint effects of smoking and PRS was 11.26 (9.99-12.68). This observed OR was less than that expected OR (13.88) under the multiplicative model and greater than the expected OR (6.91) under the additive model visualized in **Figure 3.2**. The estimated PAR for ever-smoking was 34.1% and for PRS was 60.3%. The combined PAR for smoking and PRS was 76.5%. As a continuous variable of PRS, we also observed a statistically significant sub-multiplicative interaction (OR (95% CI)=0.9957 (0.993-0.9983), P=1.16E-03) and super-additive interaction (RERI (95% CI)=0.07 (0.05-0.09), P=6.35E-25) (data not shown).

Interaction between pack-years and PRS in ever-smokers

Because we observed a statistically significant interaction between ever-smoking and PRS, we then tested whether the increased risk of COPD by smoking dose measured by pack-years was modified by PRS within ever-smokers (**Table 3.4**). The strength of the association between pack-years and COPD tended to be greater in the lowest PRS group compared to the highest PRS group. We found no evidence of statistically significant interaction on multiplicative scale (P>0.05 for all categories of pack-years). The interaction on an additive scale was statistically significant for all categories of pack-years (RERI ranged from 1.45 to 13.50, P<2.00E-03). The log scale of ORs for joint effects of pack-years and PRS are depicted in **Figure 3.3**. Estimated ORs are provided in **Supplemental Table 3.2.** The ORs (95% CI) for COPD showed an increasing trend across strata for both categories of pack-years or PRS. The estimated OR (95% CI) for joint effects of pack-years>50 and PRS (i.e. OR comparing a group of pack-years>50 in the highest PRS group to a reference group of pack-years≤10 in the lowest PRS group) was 22.11 (16.1-30.36). Of note, we observed similar ORs for COPD between groups with relatively high smoking exposure in the lowest PRS group and those with relatively low smoking exposure in the highest PRS group. For example, the OR (95% CI) for a group of pack-years>50 in the lowest PRS group was 5.79 (4.19-8.00), while for a group of 10.1<pack-

years<20 in the highest PRS group was 5.73 (4.40-7.46). Results of other smoking measures (never/former/current-smoking status and a categorical variable of pack-years with never-smoker as the reference group) are provided in **Supplemental Tables 3.3 and Supplemental Figure 3.1)**.

Discussion

We tested for possible gene-by-smoking interactions on COPD risk using a polygenic risk score (PRS) based on genetic markers achieving genome-wide significance as influencing quantitative measures lung function identified from GWAS in European population. Our study provides evidence for a sub-multiplicative interaction and super-additive interaction between smoking and PRS on the risk of COPD. Particularly, among ever-smokers, our analyses showed strong evidence for a super-additive interaction between smoking dose and PRS on risk of COPD. It indicates the expected number of cases that could be avoided by smoking prevention efforts is larger among subjects with higher genetic predisposition.

We noted a sub-multiplicative interaction between ever-smoking and PRS. Our finding is consistent with a previous report that OR (95% CI) for an interaction between PRS based on the 279 SNPs found to influence lung function and ever-smoking on risk of COPD was below one, 0.96 (0.92-0.99) (P=1.50E-02) [9]. This estimated OR suggests OR for ever-smoking is less among individuals with higher PRS (i.e. the increased risk of COPD as a result of smoking is smaller among individuals who are more genetically susceptible to COPD).

Under the additive model, however, our results showed a strong super-additive interaction where the increased risk of COPD due to smoking was greater among individuals who are more

genetically susceptible to COPD, which is consistent with our prior hypothesis. To our knowledge, this is the first report of possible gene-by-smoking interactions on risk to COPD on an additive scale using a PRS measure. In the public health context, the predicted reduction in risk of COPD from smoking intervention would be larger in individuals with a higher genetic risk. These findings may appear somewhat contradictory, but they do provide evidence for the potential application of genetic information in targeted smoking intervention programs. However, ethical issues to use genetic information in public health settings should be carefully considered.

Interestingly, we observed similar estimated ORs for COPD between groups of individuals with relatively high smoking exposure and low genetic predisposition and those of individuals with relatively low smoking exposure and high genetic predisposition among ever-smoker. Subjects in these groups were about 4-5.7 times more likely to develop COPD than subjects who had low smoking exposure and low genetic predisposition. This finding provides two important messages. First, among those who have a low genetic predisposition, there is still substantial risk of developing COPD due to smoking behavior. This highlights the importance of smoking cessation regardless of an individual's genetic susceptibility in preventing COPD. Second, for those who have a high genetic predisposition, they could reduce their risk to develop COPD by changing their smoking behavior, comparable to those who have a low genetic predisposition.

There are several limitations in our study. First, our estimated interaction measures can be inflated. We calculated RERI assuming the disease is rare in our population, but we observed a high proportion of COPD cases (>10%) in some strata of smoking and PRS. Second, our analysis was deliberately limited to two extreme groups of PRS when testing for interaction. We sought to minimize high-dimensional issue when considering 10 distinct decile groups of PRS. Considering middle groups of PRS (the vast majority of the UKB cohort) could provide a more

comprehensive picture of PRS-by-smoking interactions. Third, we did not calculate the absolute risk in this paper. We tested for additive PRS-by-smoking interaction but did not formally test for differences in actual absolute risk. Fourth, our study included only subjects of European ancestry. It is not clear whether these results can be generalized to other racial/ethnic groups.

In conclusion, our study provides strong evidence for additive interaction between smoking and PRS on the risk of COPD for the first time. Under an additive model, these findings suggest efforts of smoking intervention to prevent COPD would be larger for individuals who are more genetically predisposed to this disease.

References

1. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. *Epidemiol Method*. 2014;3(1):33-72. doi:10.1515/em-2013-0005
2. Ahmad S, Rukh G, Varga T V., et al. Gene × Physical Activity Interactions in Obesity: Combined Analysis of 111,421 Individuals of European Ancestry. *PLoS Genet*. 2013;9(7):1-9. doi:10.1371/journal.pgen.1003607
3. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. Polygenic Risk Score for Coronary Heart Disease Modifies the Elevated Risk by Cigarette Smoking for Disease Incidence. *Circ Genomic Precis Med*. 2018;11(1):e001856. doi:10.1161/CIRCGEN.117.001856
4. Garcia-Closas M, Rothman N, Figueroa JD, et al. Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Cancer Res*. 2013;73(7):2211-2220. doi:10.1158/0008-5472.CAN-12-2388
5. Aschard H, Tobin MD, Hancock DB, et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int J Epidemiol*. 2017;46(3):894-904. doi:10.1093/ije/dyw318
6. Qi Q, Chu AY, Kang JH, et al. Sugar-Sweetened Beverages and Genetic Risk of Obesity.

N Engl J Med. 2012;367(15):1387-1396. doi:10.1056/NEJMoa1203039

7. Kim W, Prokopenko D, Sakornsakolpat P, et al. Genome-wide gene-by-smoking interaction study of Chronic Obstructive Pulmonary Disease [abstract]. (Presented at the annual meeting for the American Society of Human Genetics. October 15-19, 2019, Houston, TX).

8. Hancock DB, Soler Artigas MM, Gharib SA, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet*. 2012;8(12):e1003098. doi:10.1371/journal.pgen.1003098

9. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019;51(3):481-493. doi:10.1038/s41588-018-0321-7

10. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z

11. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet*. 2019;51(3):494-505. doi:10.1038/s41588-018-0342-2

12. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am J Respir Crit Care Med*. 2017;195(5):557-582. doi:10.1164/rccm.201701-0218PP

13. Hankinson JL, Odencrantz JR, Fedan KB. *Spirometric Reference Values from a Sample of the General U.S. Population*. Vol 159.; 1999. www.atsjournals.org. Accessed September 3, 2019.

14. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology*. 1992;3(5):452-456. doi:10.1097/00001648-199209000-00012

15. García-Closas M, Jacobs K, Kraft P, Chatterjee N. Analysis of epidemiologic studies of genetic effects and gene-environment interactions. *IARC Sci Publ*. 2011;(163):281-301. http://www.ncbi.nlm.nih.gov/pubmed/22997868. Accessed November 3, 2019.

16. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population

attributable risk for multiple risk factors using case-control data. *Am J Epidemiol*. 1985;122(5):904-914. doi:10.1093/oxfordjournals.aje.a114174

Table 3.1. Subject Characteristics

	Control	COPD Case	Total
N	179689	21077	200766
Age (Mean (SD))	55.73 (8.02)	59.36 (7.31)	56.11 (8.03)
Female (%)	75348 (41.9)	10962 (52.0)	86310 (43.0)
Smoking measures			
Ever smoker (%)	59145 (32.9)	12446 (59.1)	71591 (35.7)
Former smoker (%)	49144 (27.3)	7857 (37.3)	57001 (28.4)
Current smoker (%)	10001 (5.6)	4589 (21.8)	14590 (7.3)
Pack-Years (Mean (SD))	6.09 (12.53)	19.71 (23.91)	7.52 (14.77)
Smoking duration (Mean (SD))	6.95 (12.28)	19.38 (19.08)	8.25 (13.70)
Cigarette per day (Mean (SD))	5.42 (9.66)	11.73 (12.90)	6.08 (10.23)
Polygenic risk score			
1st decile (%)	19033 (10.6)	1044 (5.0)	20077 (10.0)
10th decile (%)	16332 (9.1)	3745 (17.8)	20077 (10.0)

Table 3.2. Overall associations of smoking and lung function PRS with COPD

Exposure	OR (95% CI)
Ever smoking	
Never smoker	1.00 (Reference)
Ever smoker	2.37 (2.3-2.44)
Pack-Years	
Pack-Years	1.03 (1.03-1.03)
Category of Pack-Years	
≤10	1.00 (Reference)
10.1-20	1.55 (1.44-1.66)
20.1-30	1.65 (1.53-1.79)
30.1-40	2.99 (2.75-3.24)
40.1-50	4.20 (3.83-4.59)
>50	5.77 (5.27-6.31)
Continuous PRS	
PRS	1.04 (1.03-1.04)
Decile of PRS	
1	1.00 (Reference)
2	1.38 (1.27-1.5)
3	1.52 (1.4-1.66)
4	1.73 (1.59-1.88)
5	1.85 (1.71-2.01)
6	2.20 (2.03-2.38)
7	2.42 (2.23-2.61)
8	2.62 (2.42-2.84)
9	3.11 (2.88-3.36)
10	4.35 (4.03-4.68)

Table 3.3. Interaction between ever-smoking and PRS on the risk of COPD

	Lowest PRS group			Highest PRS group			
	n	**% of COPD case**	**OR (95% CI)**	**n**	**% of COPD case**	**OR (95% CI)**	**OR (95% CI) for PRS within smoking group**
Never	12943	0.03	1.00 (Reference)	12872	0.13	5.14 (4.56-5.79)	5.15 (4.57-5.80)
Ever	7134	0.09	2.77 (2.42-3.17)	7205	0.29	11.26 (9.99-12.68)	4.02 (3.64-4.44)
OR (95% CI) for smoking within PRS group			2.81 (2.46-3.22)			2.19 (2.02-2.37)	

Measure of interaction on multiplicative scale: ratio of ORs (95% CI)=0.79 (0.68-0.92), P=3.11E-03

Measure of interaction on additive scale: RERI (95% CI)=4.35 (3.54-5.16), P=6.23E-26

Expected OR for smoking and PRS under multiplicative model=13.88

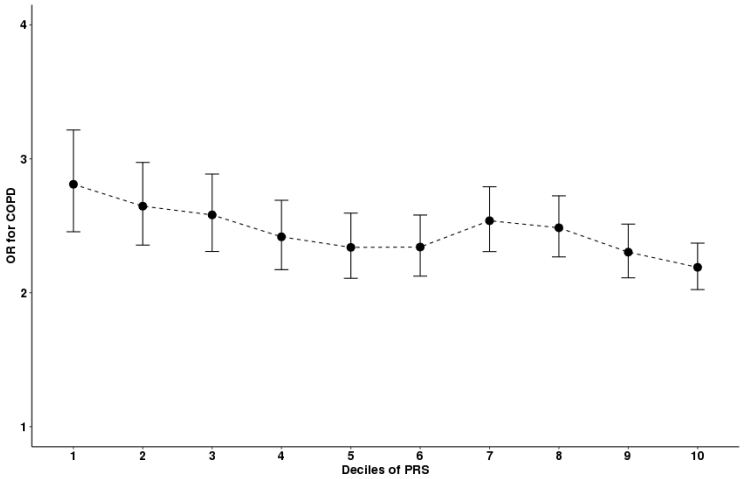
Expected OR for smoking and PRS under additive model=6.91

Table 3.4. Interaction between pack-years and PRS on the risk of COPD in ever-smokers

	Stratum-specific effects									
	Lowest PRS group			Highest PRS group			Interaction on multiplicative scale		Interaction on additive scale	
Pack-years	n	% of COPD case	OR (95% CI)	n	% of COPD case	OR (95% CI)	OR (95% CI)	P	RERI (95% CI)	P
≤10	1933	0.04	1.00 (Reference)	1769	0.13	1.00 (Reference)				
10.1-20	1886	0.05	1.30 (0.95-1.78)	1927	0.2	1.37 (1.15-1.65)	1.14 (0.8-1.64)	4.70E-01	1.45 (0.52-2.38)	2.00E-03
20.1-30	1312	0.08	1.62 (1.16-2.24)	1281	0.31	1.55 (1.27-1.9)	1.22 (0.85-1.75)	2.87E-01	2.27 (1.04-3.51)	2.95E-04
30.1-40	751	0.17	3.13 (2.25-4.35)	896	0.47	2.75 (2.22-3.4)	1.08 (0.75-1.56)	6.81E-01	6.05 (3.61-8.49)	1.17E-06
40.1-50	408	0.27	5.58 (3.94-7.89)	472	0.56	3.96 (3.08-5.08)	0.85 (0.57-1.28)	4.34E-01	10.19 (5.43-14.96)	2.79E-05
>50	458	0.32	6.66 (4.75-9.33)	471	0.65	5.11 (3.94-6.61)	0.94 (0.63-1.4)	7.76E-01	13.50 (7.44-19.56)	1.24E-05

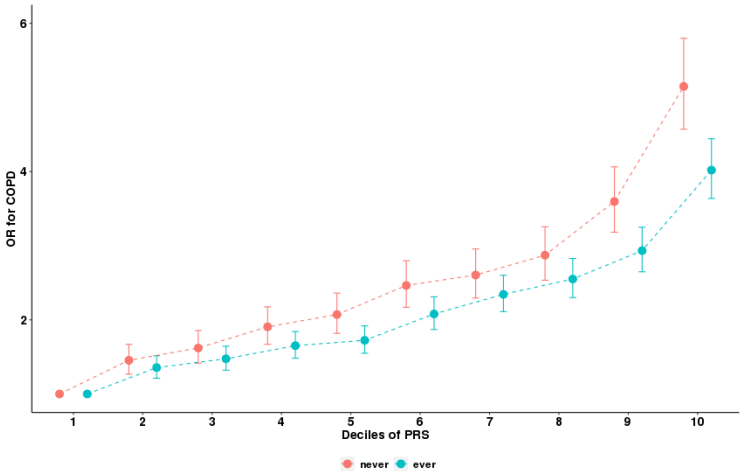
Figure 3.1. Estimated ORs for COPD stratified by deciles of a lung function PRS and by ever-smoking

Figure 3.1a. Association of ever-smoking with COPD stratified by deciles of lung function PRS



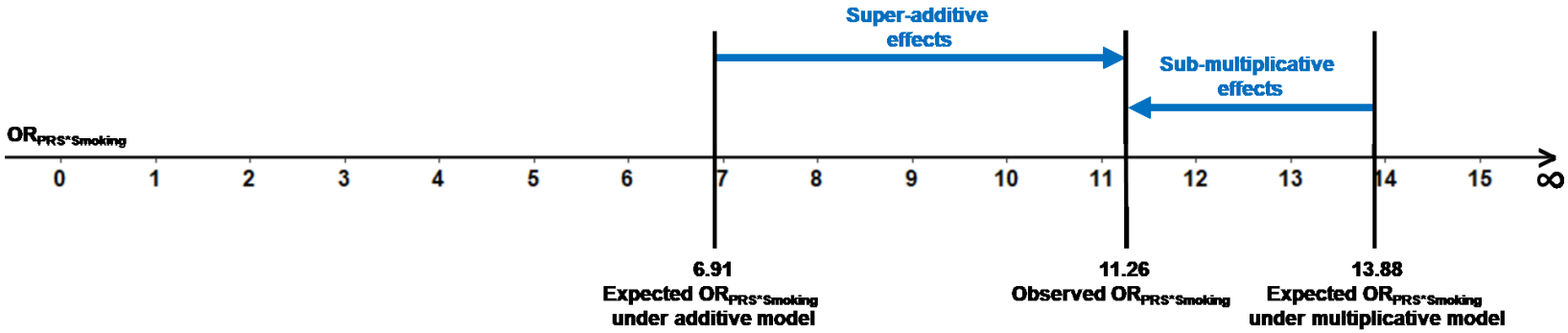
Note: Estimated ORs for COPD comparing ever-smokers to never-smokers were calculated for each decile of the lung function PRS.

Figure 3.1b. Association of lung function PRS deciles with COPD stratified by ever-smoking



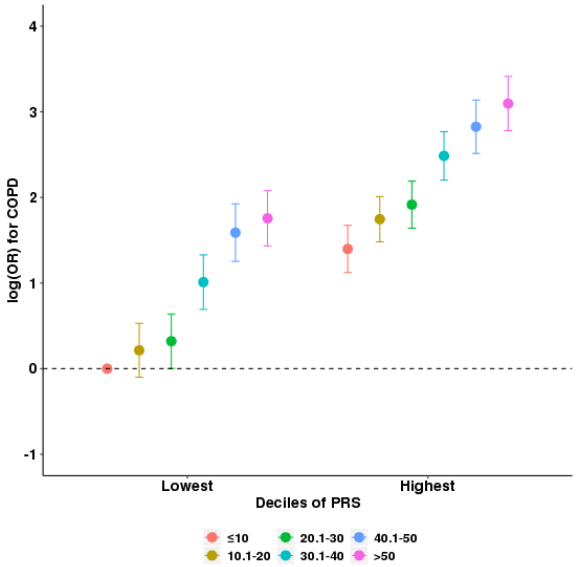
Note: Estimated ORs for COPD comparing the 2nd to 10th deciles of a lung function PRS to the 1st decile (lowest) for a lung function PRS among never- and ever- smokers.

Figure 3.2. Observed and expected ORs for interaction of ever-smoking with PRS (1st and 10th decile)



Note: A presence of sub-multiplicative interaction and super-additive interaction between ever-smoking and PRS (10^{th} vs. 1^{st} decile).

Figure 3.3. Joint effects of pack-years and a lung function PRS on COPD risk



Note: The y axis is log(OR) for COPD comparing each group to a reference group of low exposure to smoking (i.e. pack-years≤10) in the lowest PRS group. This lowest PRS group represents the 1st decile of PRS (i.e. individual with a low genetic risk), while the highest PRS group represents 10th decile of PRS (individuals with a high genetic risk).

Supplemental Table 3.1. Overall associations of smoking (Never/Former/Current and Pack-years with never smoker as reference) with COPD

Exposure	OR (95% CI)
Smoking status	
Never smoker	1.00 (Reference)
Former smoker	1.78 (1.72-1.84)
Current smoker	5.47 (5.22-5.72)
Category of Pack-Years	
Never smoker	1.00 (Reference)
≤10	1.41 (1.33-1.51)
10.1-20	2.01 (1.91-2.12)
20.1-30	1.82 (1.72-1.92)
30.1-40	3.34 (3.16-3.54)
40.1-50	4.76 (4.44-5.11)
>50	6.57 (6.13-7.04)

Supplemental Table 3. 2. Interaction between pack-years and PRS on the risk of COPD in ever-smokers

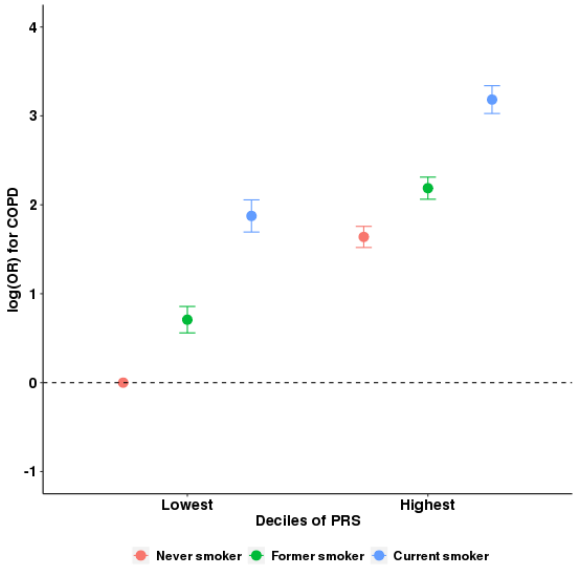
	Joint effects					
	Lowest PRS group			Highest PRS group		
Pack-years	n	% of COPD case	OR (95% CI)	n	% of COPD case	OR (95% CI)
≤10	1933	0.04	1.00 (Reference)	1769	0.13	4.05 (3.07-5.33)
10.1-20	1886	0.05	1.24 (0.9-1.7)	1927	0.2	5.73 (4.40-7.46)
20.1-30	1312	0.08	1.38 (1.00-1.89)	1281	0.31	6.79 (5.16-8.94)
30.1-40	751	0.17	2.75 (2-3.77)	896	0.47	12 (9.05-15.92)
40.1-50	408	0.27	4.9 (3.5-6.85)	472	0.56	16.86 (12.35-23.01)
>50	458	0.32	5.79 (4.19-8.00)	471	0.65	22.11 (16.10-30.36)

Supplemental Table 3.3. Interaction between smoking (Never/Former/Current and Pack-years with never smoker as reference) and PRS on the risk of COPD

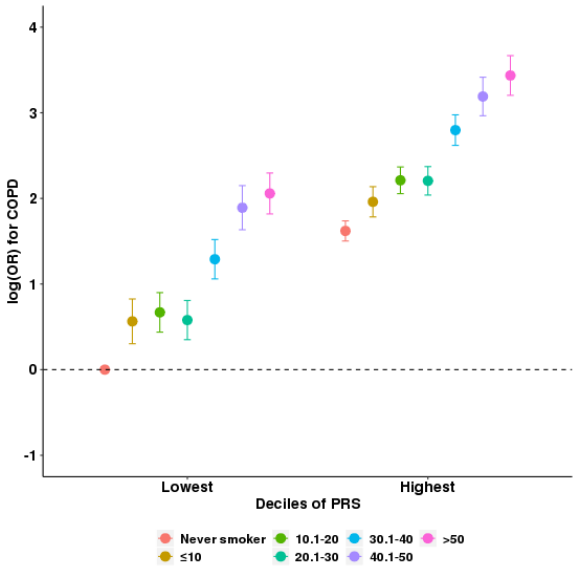
	Stratum-specific effects										Joint effects	
	Lowest PRS group			Highest PRS group			Interaction on multiplicative scale		Interaction on additive scale		Lowest PRS group	Highest PRS group
	n	% of COPD case	OR (95% CI)	n	% of COPD case	OR (95% CI)	OR (95% CI)	P	RERI (95% CI)	P	OR (95% CI)	OR (95% CI)
Smoking status												
Never smoker	12943	0.03	1.00 (Reference)	12872	0.13	1.00 (Reference)					1.00 (Reference)	5.15 (4.57-5.8)
Former smoker	5743	0.07	2.02 (1.74-2.35)	5689	0.24	1.75 (1.6-1.91)	0.85 (0.72-1.01)	6.69E-02	2.86 (2.14-3.59)	1.08E-14	2.03 (1.75-2.36)	8.91 (7.87-10.09)
Current smoker	1391	0.19	7.29 (6.06-8.78)	1516	0.45	4.52 (3.97-5.14)	0.72 (0.58-0.9)	3.18E-03	13.00 (9.85-16.15)	5.97E-16	6.52 (5.44-7.81)	24.13 (20.64-28.21)
Pack-Years												
Never smoker	12943	0.03	1.00 (Reference)	12872	0.13	1.00 (Reference)					1.00 (Reference)	5.05 (4.49-5.69)
≤10	1933	0.04	1.57 (1.21-2.05)	1769	0.13	1.46 (1.25-1.7)	0.8 (0.59-1.08)	1.43E-01	2.06 (0.75-3.38)	2.00E-03	1.76 (1.35-2.28)	7.1 (5.95-8.47)
10.1-20	1886	0.05	1.92 (1.53-2.42)	1927	0.2	1.82 (1.6-2.07)	0.92 (0.71-1.2)	5.61E-01	3.46 (2.20-4.72)	7.87E-08	1.95 (1.55-2.46)	9.13 (7.81-10.67)
20.1-30	1312	0.08	2.01 (1.59-2.54)	1281	0.31	1.73 (1.5-2)	1.01 (0.77-1.31)	9.65E-01	2.86 (1.63-4.09)	5.36E-06	1.78 (1.42-2.24)	9.07 (7.68-10.71)
30.1-40	751	0.17	3.97 (3.15-5.01)	896	0.47	3.17 (2.71-3.71)	0.89 (0.68-1.18)	4.23E-01	8.50 (5.88-11.13)	2.02E-10	3.63 (2.89-4.57)	16.4 (13.72-19.61)
40.1-50	408	0.27	7.07 (5.46-9.15)	472	0.56	4.8 (3.89-5.92)	0.73 (0.52-1.01)	5.56E-02	14.39 (8.85-19.93)	3.58E-07	6.63 (5.12-8.58)	24.3 (19.4-30.43)
>50	458	0.32	8.43 (6.62-10.74)	471	0.65	6.06 (4.88-7.53)	0.78 (0.57-1.08)	1.36E-01	19.88 (12.64-27.11)	7.45E-08	7.83 (6.16-9.94)	31.04 (24.62-39.15)

Supplemental Figure 3.1. Joint effects of smoking measures and a lung function PRS on COPD risk

Supplemental Figure 3.1a. Smoking status



Supplemental Figure 3.1b. Pack-years



Note: The y axis is log(OR) for COPD comparing each group to a reference group of never-smoker with the lowest PRS (low genetic risk, i.e. the 1st decile of PRS). The highest PRS group represents 10th decile of PRS (high genetic risk).

Chapter 5. Summary of key findings and conclusions

Chapter 5. Summary of key findings and conclusion

Chronic Obstructive Pulmonary Disease (COPD) is the third leading cause of death worldwide. Although cigarette smoking is the most important environmental risk factor, risk of COPD is influenced by the interplay between genetic and environmental risk factors. COPD is highly heterogenous with two main subtypes, emphysema and small airway disease, which can be quantitatively measured by computed tomography (CT) imaging. Most previous research has relied on identifying associations of common variants with quantitative measures of emphysema in cross-sectional studies. The adverse effects of smoking on risk of COPD vary by an individual's susceptibility, raising the possibility of gene-by-smoking interaction. However, little is known about gene-by-smoking interactions on COPD risk to date. This dissertation addresses these issues with the following specific aims:

Aim 1. To examine genetic variants for association with change in quantitative emphysema measured by CT imaging from two longitudinal cohort studies: COPDGene and Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE).

Aim 2. To identify novel genetic variants for risk of COPD while accounting for potential smoking interactions and assess the gene-by-smoking interactions on risk of COPD at known COPD and lung function GWAS loci.

Aim 3. To examine whether effects of smoking on COPD risk are modified by a polygenic risk score of lung function.

Summary of key findings

Analysis of longitudinal change in quantitative emphysema. We conducted a GWAS to identify variants associated with annual change in quantitative emphysema measured by CT scans and examined effects of variants previously associated with cross-sectional quantitative emphysema, COPD and spirometric measures of lung function on the annual change in emphysema from two large cohorts, COPDGene and ECLIPSE. None of single nucleotide polymorphisms (SNPs) yielded genome-wide significance. However, in our candidate region analysis, we identified significant associations for a variant in *DSP* and a polygenic risk score (PRS) based on spirometric measures of lung function with the annual change in emphysema in individuals of European ancestry. This study was published on the "Respiratory Research" journal in 2019 [1].

Analysis of SNP-by-smoking interaction on COPD. We conducted a genome-wide gene-by-smoking interaction study of risk to COPD, using a 2df joint test for genetic main effects and gene-by-smoking interaction effects on COPD risk. Most of the significant signals from the 2df test had been reported in previous GWAS of COPD. However, we identified two loci, 16q22.1 – SMPD3 and 19q13.2 – EGLN2, reaching genome-wide significance. We detected a genome-wide significant interaction with smoking at the 15q25.1 region which contains – CHRNB4, previously identified in studies of COPD and smoking behavior. We confirmed *SERPINA1* PI*Z-by-smoking interaction on risk to COPD.

Analysis of polygenic risk score-by-smoking interaction on risk to COPD. We focused on tests of gene-by-smoking interaction on COPD risk using a risk score based on 279 genetic risk factors for lung function identified from a previous GWAS of lung function in the UK Biobank study. Our study provides evidence for interaction between smoking and PRS on the risk of COPD. The

results were most compelling under an additive model. Particularly, among ever-smokers, our analyses showed strong evidence for a super-additive interaction between smoking dose and PRS on the risk of COPD.

Strengths and limitations

Strengths. This dissertation has several key strengths. First, the COPDGene and ECLIPSE studies are the largest cohorts to date to measure emphysema using chest CT scans. CT measures were carefully quantified, allowing to assess the progression of emphysema and its genetic factors. Second, this dissertation included large sample size over 200,000 subjects from the UK Biobank (UKB), a large-scale cohort recruiting volunteers throughout UK, collecting a wide range of phenotypes as well as genetic data. Third, this is the first GWAS investigating change in quantitative emphysema measured by CT imaging in longitudinal data. This provides the first evidence of associations of a *DSP* variant and a lung function PRS with longitudinal emphysema progression measured by CT scan. Fourth, this is one of the first studies to investigate gene-by-smoking interaction at the genome-wide scale on risk to COPD itself. Fifth, this study provides evidence of additive interaction between smoking and a lung function PRS on the risk of COPD for the first time.

Limitations. This dissertation should be interpreted under several limitations. First, the follow-up period in our study of quantitative measures of emphysema may be too short to capture the natural history of emphysema in adults. Second, we only included independent subjects of European ancestry to test for interactions using the UK Biobank data. Third, the additive interaction measures can be exaggerated in analysis of PRS-by-smoking interaction on risk to COPD. We assessed the additive interaction using Relative Excess Risk due to Interaction (RERI) measure assuming the disease was rare in our population, but the high prevalence of

COPD (>10%) in the sub-groups for the stratified analysis of PRS and smoking groups may undermine this assumption. Fourth, we did not calculate the actual absolute risk and evaluate the risk differences when estimating the additive interaction between smoking and a lung function PRS. Fifth, a "healthy volunteer" selection bias exists in the UKB study. The UKB cohort is not representative of the general population; UKB participants are less likely to smoke and have fewer self-reported health conditions compared with the general population of the UK [2]. However, generalizability is not necessary to make an inference of associations. Its large sample size and heterogeneity of smoking exposures would still make our findings valid.

Future directions

Additional investigation of the DSP gene. We identified an association between rs2076295 in the *DSP* gene and longitudinal change in quantitative emphysema. This variant is also associated with increased risk of COPD [3], reduced lung function [4], as well as decreased risk of pulmonary fibrosis [5]. Whether this association represents true progression of emphysema or (for the opposite allele) development of fibrosis, or both needs confirmation in further studies.

Ongoing longitudinal follow-up. Our follow-up period (approximately 5.5 years in COPDGene and 2.6 years in ECLIPSE) is sufficient to show emphysema progression, but very short in the context of the natural history of emphysema. Longer follow-up may provide more accurate data to estimate the emphysema progression.

Diversity of population. Investigation of more ethnically diverse populations should lead to more robust inferences about gene-by-environment interactions by increasing diversity of not only environmental exposure but also genetic exposure [6].

Integration with "-omics" data. Integration of genetic markers and other "-omics" data (transcriptomic, proteomic or epigenomic data) would be helpful to understand gene-by-smoking interactions on COPD risk or more broadly, gene-by-environment interactions. For example, genetic markers influencing other biomarkers such as expression Quantitative Trait Loci (eQTL) may be more likely to interact with smoking [7].

Evaluation of additive interaction. Additive interactions are less explored than multiplicative interactions in gene-by-environment interaction studies, although it has more public health relevance. Our finding highlights the importance of assessing the additive interaction as well as the multiplicative interaction in future gene-by-environment interaction studies.

Public health significances

This dissertation has public health significances. Particularly, the evidence of super-additive interaction between smoking and PRS indicates reduction in risk of COPD by smoking cessation programs should be larger among individuals at higher genetic risk of COPD. Our findings provide proof-of-principle for the potential application of genetic information in targeting smoking intervention programs. However, ethical issues to use genetic information in public health settings should be carefully considered. Improved understanding of the genetics and potential gene-by-environment interactions influencing risk to COPD may aid in the discovery of novel therapeutics and a more tailored treatment, and ultimately reduce the burden of COPD.

References

1. Kim W, Cho MH, Sakornsakolpat P, et al. DSP variants may be associated with longitudinal change in quantitative emphysema. *Respir Res*. 2019;20(1):160. doi:10.1186/s12931-019-1097-8

2. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *Am J Epidemiol*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246
3. Hobbs BD, de Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
4. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019;51(3):481-493. doi:10.1038/s41588-018-0321-7
5. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet*. 2013;45(6):613-620. doi:10.1038/ng.2609
6. Ritz BR, Chatterjee N, Garcia-Closas M, et al. Lessons Learned From Past Gene-Environment Interaction Successes. *Am J Epidemiol*. 2017;186(7):778-786. doi:10.1093/aje/kwx230
7. Favé M-J, Lamaze FC, Soave D, et al. Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nat Commun*. 2018;9(1):827. doi:10.1038/s41467-018-03202-2

Curriculum Vitae

Name: Woori Kim

Email: wkim42@jhu.edu; woori.kim11@gmail.com

Office Address: 615 N. Wolfe Street, Baltimore, MD, 21202

Home Address: 111 W. Centre Street, Baltimore, MD, 21201

Work Phone: 1-410-955-3474

Personal Phone: 1-443-467-3305

Education

Aug. 2015 – Jan. 2020	Doctor of Philosophy, Department of Epidemiology, Johns Hopkins School of Public Health, MD, U.S. Dissertation: Genome-wide Study of Emphysema Progression and Gene-by-Smoking Interaction of Chronic Obstructive Pulmonary Disease
Mar. 2012 - Feb. 2014	Master of Public Health, Graduate School of Public Health, Seoul National University, South Korea Thesis: Genetic and Environmental Effects on Myopia: The Healthy Twin Study
Mar. 2007 - Feb. 2012	Bachelor of Public Health, Department of Environmental Health, College of Health Science, Korea University, South Korea

Research Experience

June 2019 – Aug. 2019	Internship	Biogen, Statistical Genetics & Genetic Epidemiology and Human Target Validation Core, Boston, MA, Dr. Paola Bronson
Sep. 2018 – May 2019	Research Trainee	Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, Drs. Edwin K. Silverman and Michael H. Cho
June 2016 - Current	Research Assistant	Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore MD, Dr. Terri H. Beaty

June 2016 – Aug. 2017	Research Assistant	Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore MD, Dr. Christine Ladd-Acosta
Oct. 2014 – Aug. 2015	Researcher	Molecular Epidemiology Branch, Division of Cancer Epidemiology and Prevention, National Cancer Center, South Korea, Dr. Jeongseon Kim
Mar. 2014 – Sep. 2014	Researcher	Graduate School of Public Health, Seoul National University, South Korea, Dr. Joohon Sung
Mar. 2012 – Feb. 2014	Research Assistant	Graduate School of Public Health, Seoul National University, South Korea, Dr. Joohon Sung
June 2011 – Nov. 2011	Undergraduate Internship	College of Health Science, Korea University, South Korea, Dr. Jong-Tae Lee
Jan. 2010	Research Assistant	Department of Food and Drug, Seoul Metropolitan Government Research Institute of Public Health and Environment, South Korea

Publications

Kim W, Prokopenko D, Sakornsakolpat P, Hobbs BD, Lutz SM, Hokanson JE, Wain LV, Melbourne CA, Shrine N, Tobin MD, Silverman EK, Cho MH, Beaty TH. Genome-wide gene-by-smoking interaction study of Chronic Obstructive Pulmonary Disease. 2019 American Journal of Epidemiology. Under review

Kim W, Cho MH, Sakornsakolpat P, Lynch DA, Coxson HO, Tal-Singer R, Silverman EK, Beaty TH. DSP variants may be associated with longitudinal change in quantitative emphysema. Respiratory research. 2019 Dec;20(1):160.

Kim W, Woo HD, Lee J, Choi IJ, Kim YW, Sung J, Kim J. Dietary folate, one-carbon metabolism-related genes, and gastric cancer risk in Korea. Molecular nutrition & food research. 2016 Feb;60(2):337-45.

Lim DH, **Kim W**, Han G, Bae GH, Kim MH, Kim MJ, Song YM, Chung ES, Chung TY, Sung J. The heritability of corneal and ocular higher-order aberrations in Koreans: The Healthy Twin Study. Investigative ophthalmology & visual science. 2015 Jun 1;56(6):3919-28. **(**Lim DH and **Kim W** are co-first authors.)

Kim MH, Zhao D, **Kim W**, Lim DH, Song YM, Guallar E, Cho J, Sung J, Chung ES, Chung TY. Heritability of myopia and ocular biometrics in Koreans: the healthy twin study. Investigative ophthalmology & visual science. 2013 May 1;54(5):3644-9.

Presentations

Oct. 2019	Genome-wide gene-by-smoking interaction study of chronic obstructive pulmonary disease, Poster session, American Society of Human Genetics Annual Meeting

May 2019	*DSP* variants may be associated with longitudinal change in quantitative emphysema, Poster session, American Thoracic Society International Conference
April 2019	Genome-wide gene-smoking interaction on COPD in UK Biobank study, Poster session, Genetics Research Day, The Maryland Genetics, Epidemiology, and Medicine Training Program, Johns Hopkins School of Public Health and School of Medicine
Oct. 2018	Genome-wide association study of longitudinal change in quantitative emphysema, Poster session, American Society of Human Genetics Annual Meeting
June 2018	Classification of sub-groups of COPD and emphysema cases defined on CT scan and spirometry: COPDGene study, Poster session, Society for Epidemiologic Research Annual Meeting
Feb. 2018	Genome-wide association study of emphysema progression, Poster session, Genetics Research Day, The Maryland Genetics, Epidemiology, and Medicine Training Program, Johns Hopkins School of Public Health and School of Medicine
Feb. 2017	Elucidation of relationship among chronic obstructive pulmonary disease, genes and smoking, Poster session, Genetics Research Day, The Maryland Genetics, Epidemiology, and Medicine Training Program, Johns Hopkins School of Public Health and School of Medicine
Nov. 2015	Dietary folate, one-carbon metabolism-related genes, and gastric cancer risk in Korea, Centennial Poster Session, Department of Epidemiology, Johns Hopkins School of Public Health
Nov. 2012	Genetic variants on 19q13.12 are associated with refractive error in Korean populations, Poster session, American Society of Human Genetics Annual Meeting

Professional Development

July 2018	Human and mammalian genetics and genomics: the 59th Mckusick Short Course
June 2018	Teaching Assistant, Intermediate Epidemiology
Dec. 2017	Grant Application, Lung Health Dissertation Grant, American Lung Association
Oct. 2017 – Dec. 2017	Teaching Assistant, Principle of Genetic Epidemiology II
Sep. 2017 – Oct. 2017	Teaching Assistant, Epidemiologic Inference in Public Health I
Jan. 2017 – Mar 2017	Teaching Assistant, Epidemiologic Methods 3
Sep. 2016 – May 2017	Genetic Epidemiology Journal Club Coordinator

Sep. 2016 – Oct. 2016	Teaching Assistant, Epidemiologic Inference in Public Health I

Honor/Award

Sep. 2018 – May 2019	COPDGene Research Fellowship, designed for trainees and early-stage investigators to visit a specialized center from any COPDGene cohort centers, Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, PI: Dr. Edwin K. Silverman, Mentor: Dr. Michael H. Cho
April 2019	Basic Science Prize for the abstract entitled with "*DSP* variants may be associated with longitudinal change in quantitative emphysema", The 74th Annual Meeting of the Massachusetts Pulmonary Section
Nov. 2017	Mogam Scholarship, Mogam Science Scholarship Foundation, South Korea
Sep. 2017 – Aug. 2018	Research Associate, The Maryland Genetics, Epidemiology, and Medicine Training Program, Johns Hopkins School of Public Health and School of Medicine, Baltimore, MD
May 2017	The Ellen B. Gold Fund for Epidemiology, Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD
Sep. 2015 – Dec. 2019	Merit-based Tuition Scholarship, Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD
Spring 2013	Merit-based Scholarship, Seoul National University, South Korea
Fall 2013	Merit-based Scholarship, Seoul National University, South Korea
Fall 2012	Merit-based Scholarship, Seoul National University, South Korea
Oct. 2011	Excellence Award, Korean Society of Environmental Health, Undergraduate student oral presentation session, Title: Children Health and Climate Change, South Korea
Spring 2010	Spring Honors Scholarship, Korea University, South Korea